DOCUMENT RESUME

ED 340 714 TM 017 386

TITLE	TOEFL Test and Score Manual, 1990-91.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	90
NOTE	44p.; An errata slip, following the cover, corrects Tables 9 and 10.
AVAILABLE FROM	Test of English as a Foreign Language, P.O. Box 6155, Princeton, NJ 08541-6155.
PUB TYPE	Guides - Non-Classroom Use (055) -- Reports - Descriptive (141)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	College Entrance Examinations; *English (Second Language); *Essay Tests; Expressive Language; Foreign Students; Higher Education; *Listening Comprehension Tests; *Reading Comprehension; Scoring; Student Placement; Testing Programs; Test Interpretation; Test Manuals; Vocabulary Skills; *Writing Tests
IDENTIFIERS	*English Placement Tests; *Test of English as a Foreign Language; Test of Written English

ABSTRACT

This manual has been prepared for those responsible for interpreting scores on the Test of English as a Foreign Language (TOEFL). In addition to test interpretation information, the manual describes the test, explains the TOEFL program, and discusses program research activities. The TOEFL was developed in 1963 to test the English-language proficiency of those wanting to study at colleges and universities in the United States. The TOEFL consists of the following sections: (1) listening comprehension; (2) structure and written expression; and (3) vocabulary and reading comprehension. A direct assessment of writing proficiency, the Test of Written English, is a required section of the TOEFL; it is a 30-minute essay test that demonstrates a student's ability to organize ideas on paper. Test questions are written by language specialists, members of the TOEFL Committee of Examiners, and test specialists with the Educational Testing Service (ETS), which administers the TOEFL program. The TOEFL program provides 12 possible test dates each year, although actual administrations vary in a given area. Tests and scores are reported by ETS. Institutions can use the TOEFL for internal placement and assessment purposes. Following an overview of the program, this manual contains information on testing; test results; use of TOEFL scores; performance of reference groups; statistical characteristics of the test; and other TOEFL programs, procedures, publications, and services. A 34-item list of references and 19 data tables are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340714

1990-91

TOEFL

Test and Score Manual

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

ANNE F. MARCANTONIO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

25

CELEBRATING 25 YEARS OF SERVICE

TM017386

BEST COPY AVAILABLE

****

ERRATA NOTICE

****

Please note corrections to Tables 9 and 10 on pages 23-24 of the 1990-91 TOEFL Test and Score Manual.

Table 9

Language	Total Score Mean Listed	Corrected Total Score Mean
Arabic	480	482
Chinese	509	516
Japanese	485	491
Korean	505	509
Spanish	534	536

Table 10

Country	Total Score Mean Listed	Corrected Total Score Mean
China, People's Republic of	521	525
Hong Kong	506	510
India	571	573
Japan	485	491
Korea	505	509
Taiwan	505	509

Copyright © 1990 Educational Testing Service, Princeton, NJ. All rights reserved.

TOEFL
Test and Score Manual

The TOEFL® *Test and Score Manual* has been prepared for deans, admissions officers and graduate department faculty, administrators of scholarship programs, ESL teachers, foreign student advisers, and others responsible for interpreting TOEFL test scores. In addition to providing information about the interpretation of TOEFL scores, the *Manual* describes the test, explains the operation of the TOEFL program, and discusses program research activities.

This edition of the *Test and Score Manual* updates material in the 1987 edition, providing more recent statistical data and other information of interest to score users. It incorporates ideas and suggestions offered by users of the *Manual*, the staff of Educational Testing Service (ETS), and others associated with the program.

The TOEFL program generally publishes a new edition of the *Manual* every two years. The edition in use at the time a score report is released is indicated on the score report.

Guidelines for the Use of TOEFL Scores, an abbreviated version of the *Manual*, contains selected information about the test and about the interpretation of score results.

Your suggestions for improving the usefulness of the TOEFL *Test and Score Manual* will be most welcome. Please feel free to send any comments to us.

Director
TOEFL Programs & Services
P.O. Box 6155
Princeton, NJ 08541-6155
USA

Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.
Copyright © 1990 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, GRE, and TOEFL are trademarks of Educational Testing Service registered in the U.S.A. and in many other countries. Educational Testing Service, Graduate Record Examinations, NTE, SLEP, SPEAK, TSE, and TOEFL are U.S. registered trademarks.

Advanced Placement Program, College-Level Examination Program, SAT, and College Board are registered trademarks of the College Entrance Examination Board.

No part of this book may be reproduced, in any form or by any means, without permission in writing from Educational Testing Service, Princeton, NJ, USA.

Contents

Overview of the TOEFL Program 5

TOEFL Policy Council 5

TOEFL Score Recipients 6

What TOEFL Is Like 7

Background 7
The Three-Section Test 7
Test of Written English 7
Development of TOEFL Test Questions 8

The TOEFL Testing Programs 8

Friday and Saturday Testing Programs 8
Institutional Testing Program 9

TOEFL Test Results 9

Examinee Score Records 9
Additional Score Reports 10
Test Score Data Retention 10
Confidentiality of Information 10
Contents of TOEFL Score Reports 10
TOEFL Scores 11
TWE Scores 13
Hand-Scoring Service 13
Scores of Questionable Validity 13

Use of TOEFL Test Scores 14

Who Should Take TOEFL? 14
Guidelines for Using TOEFL Test Scores 15
Examinees with Disabilities 17
Appropriate Use of TOEFL Scores 17

Performance of Reference Groups on the TOEFL Test 20

Statistical Characteristics of the Test 22

Level of Difficulty 22
Adequacy of Time Allowed 22
Reliabilities and the Standard Error of Measurement 25
Reliability of Gain Scores 25
Intercorrelations Among Scores 26
Validity 26

Other TOEFL Programs 30

Test of Written English 30
Test of Spoken English 31
Speaking Proficiency English Assessment Kit 31
Secondary Level English Proficiency Test 32

Procedures at Test Centers 32

Measures to Protect Test Security 32

TOEFL Publications and Services 34

TOEFL Research Reports 36

References 38

ETS Field Service Offices 39

TOEFL Representatives 40

Overview of the TOEFL Program

The Test of English as a Foreign Language (TOEFL) was developed in 1963 through the cooperative effort of more than thirty organizations, public and private. A National Council on the Testing of English as a Foreign Language was formed, composed of representatives of private organizations and government agencies concerned with testing the English proficiency of nonnative speakers of English who wished to study at colleges and universities in the United States. The program was financed by grants from the Ford and Danforth Foundations and was, at first, attached administratively to the Modern Language Association. In 1965, the College Board and Educational Testing Service (ETS) assumed joint responsibility for the program.

In recognition of the fact that many who take the TOEFL test are potential graduate students, a cooperative arrangement for the operation of the program was entered into by Educational Testing Service, the College Board, and the Graduate Record Examinations (GRE) Board in 1973. Under this arrangement, ETS is responsible for administering the TOEFL program according to policies determined by a fifteen-member Policy Council.

Educational Testing Service. ETS is a nonprofit organization committed to the development and administration of responsible testing programs, the creation of advisory and instructional services, and research on techniques and uses of measurement, human learning and behavior, and educational development and policy formation. It develops and administers tests, registers examinees, and operates test centers for various sponsors. ETS also supplies related services; e.g., it scores tests; records, stores, and reports test results; performs validity studies and other statistical studies; and undertakes program research. All ETS activities are governed by a sixteen-member board of trustees composed of persons from the fields of education and public service.

In addition to the College Board testing programs, the Test of English as a Foreign Language, and the Graduate Record Examinations, ETS develops and administers a number of other tests, including the Secondary School Admission Test, the Graduate Management Admission Test, and NTE* Programs tests.

The Center for Occupational and Professional Assessment (COPA) at ETS develops programs and services related to the assessment of occupational and professional competency.

College Board. The College Board is an independent nonprofit organization with a membership of more than 2,300 colleges and universities, schools, and educational associations and agencies. Representatives of the members serve on the board of trustees and on advisory councils and committees that review the programs of the College Board and participate in the determination of its policies and activities.

The College Board sponsors tests, publications, software, and professional conferences and training in the areas of guidance, admissions, financial aid, credit by examination, and curriculum improvement in order to increase student access to higher education. It also supports and publishes research studies about tests and measurement and conducts studies on education policy developments, financial aid need assessment, admissions planning, and related education management topics.

One major College Board service, the Admissions Testing Program, includes the SAT with the multiple-choice Test of Standard Written English (TSWE), and fourteen subject-matter Achievement Tests. The College Board contracts with ETS to develop these tests, operate test centers in the United States and other countries, score the answer sheets, and send score reports to examinees and to the institutions they designate as recipients.

Graduate Record Examinations Board. The GRE Board is an independent board affiliated with the Association of Graduate Schools and the Council of Graduate Schools in the United States. It is composed of eighteen members, most of whom are associated with graduate education. Standing committees of the board include the Research Committee, the Services Committee, and the Minority Graduate Education Committee.

ETS carries out the policies of the GRE Board and, under the auspices of the board, administers and operates the GRE program. Two types of tests are offered: a General Test and Subject Tests in fifteen disciplines. ETS develops the tests with faculty committees, operates test centers in the United States and other countries, scores the answer sheets, and sends score reports to the examinees and to the institutions and fellowship sponsors the examinees designate as recipients. ETS also provides information, technical advice, and professional counsel, and it develops proposals to achieve the goals formulated by the board.

In addition to its tests, the GRE program offers research, publications, and advisory services to assist graduate schools and departments in admission, guidance and placement, and with the selection of fellowship recipients.

TOEFL Policy Council

Policies governing the TOEFL program are formulated by a Policy Council of fifteen members. The College Board and the GRE Board each appoint three members to the Council. These six members comprise the Executive Committee and elect the remaining nine members, at least one of whom represents Canadian interests. Some of these members-at-large are affiliated with such institutions and agencies as graduate schools, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government. Others are specialists in the field of English as a foreign or second language.

At the present time there are four standing committees of the Council, each responsible for specific areas of program activity.

Committee of Examiners

The TOEFL Committee of Examiners is composed of six U.S. and Canadian specialists in linguistics, language testing, or the teaching of English as a foreign or second language. Members are appointed by the Policy Council and are rotated on a regular basis to ensure the continued introduction of new ideas and philosophies related to second language teaching and testing.

The primary responsibility of this committee is to establish overall guidelines for the test content, thus assuring that TOEFL is a valid measure of English language proficiency reflecting the most current trends and methodologies in the field. The committee determines the skills to be tested, the kinds of questions to be asked, and the appropriateness of the test in terms of subject matter and cultural content. Committee members review and approve the policies and specifications that govern the test content.

The Committee of Examiners not only lends its own expertise to the test and the test development process but on occasion invites the collaboration of other authorities in the field, through invitational conferences and other activities, to contribute to the improvement of the test. The committee works with ETS test development specialists in the actual development and review of test materials.

Research Committee

An ongoing program of research related to the TOEFL test is carried out under the direction of the Research Committee. Its six members include representatives of the Policy Council and the Committee of Examiners, as well as outside U.S. and Canadian specialists from the academic community who are appointed by the Policy Council. The committee reviews and approves proposals for test-related research and sets guidelines for the entire scope of the TOEFL research program.

Because the studies involved are specific to the test and the testing program, most of the actual research work is conducted by ETS staff members rather than by outside researchers. However, many projects require the cooperation of consultants and other institutions, particularly those with programs in the teaching of English as a foreign or second language. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL office.

As research studies are completed, reports are published that are available to anyone interested in the test. A list of those in print at the time this *Manual* was published is given on page 36.

Services Committee

This six-member committee of Policy Council members is responsible for reviewing and making recommendations for improving and modifying existing program services; initiating proposals for the development of new program services; and executing additional tasks requested by the Executive Committee or the Council.

Test of Written English (TWE) Core Reader Group

This seven-member group, appointed by the Executive Committee of the Policy Council, consists of U.S. and Canadian writing and ESL composition specialists with expertise in writing assessment.

The Core Reader Group, with ETS test development specialists, is responsible for developing, reviewing, and approving test items for the TWE test. They also prepare item writer guidelines and may suggest research or make recommendations for improving TWE to ensure that the test is a valid measure of academic writing.

TOEFL Score Recipients

The purpose of the Test of English as a Foreign Language is to evaluate the English proficiency of people whose native language is not English. The test was initially developed to measure the English proficiency of international students wishing to study at colleges and universities in the United States and Canada, and this continues to be its primary function. However, a number of academic institutions in other countries, as well as certain independent organizations, agencies, and foreign governments, have also found the test scores useful. The TOEFL test is recommended for students at the eleventh-grade level or above; the test content is considered inappropriate for younger students.

The TOEFL program encourages use of the test scores by an institution or organization if such use will assist it in making valid decisions, in terms of its own requirements, concerning English language proficiency. However, the institution or organization itself must determine whether TOEFL is an appropriate test, with respect to both the language skills it measures and its level of difficulty, and must establish its own levels of acceptable performance on the test.

TOEFL score users are invited to consult with the TOEFL program staff about their current or intended uses of the test results. The TOEFL office will assist institutions and organizations contemplating use of the test by providing information about its applicability and validity in particular situations. It also will investigate complaints or information about questionable interpretation or use of reported TOEFL test scores.

What TOEFL Is Like

Background

The TOEFL test originally contained five sections. As a result of extensive research (Pike, 1979; Pitcher and Ra, 1967; Swineford, 1971; *Test of English as a Foreign Language: Interpretive Information*, 1970), a three-section test was developed. The three-section test was introduced in 1976, and, by 1979, it was used in all TOEFL programs.

The Three-Section Test

Each form of the current TOEFL test consists of three separately timed sections; the questions in each section are in multiple-choice format with four possible answers or options per question. All responses are gridded on answer sheets that are computer scored. The total test time is approximately two and one-half hours; however, approximately three and one-half hours are needed for a test administration in order to admit examinees to the testing room, allow them to enter identifying information on their answer sheets, and distribute and collect the test materials. Brief descriptions of the three sections of the test follow.

- Section 1, Listening Comprehension, measures the ability to understand English as it is spoken in North America. The oral features of the language are stressed. The problems tested include vocabulary and structures that are most frequently used in spoken English, and sound and intonation distinctions that have proven to be difficult for nonnative speakers. The stimulus material and oral questions are recorded in standard North American English; the response options are printed in the test books.

  There are three parts in the Listening Comprehension section, each of which contains a specific type of comprehension task. The first part requires the examinee to choose one printed statement that is closest in meaning to a statement heard on the recording. The second part consists of a number of short conversations between two speakers, each followed by a single spoken question. The examinee must choose the best response to the question about the conversation from the four options printed in the test book. In the third part of this section, the examinee hears either a long conversation or a short talk or lecture. The talks are about a variety of subjects. After each conversation or lecture the examinee is asked three to five questions about what was heard and, for each, must choose the one best answer from the choices in the test book. Questions for all parts are spoken only one time.

- Section 2, Structure and Written Expression, measures mastery of important structural and grammatical points in standard written English.

  The language tested is formal, rather than conversational. The topics of the sentences are of a general academic nature so that individuals in specific fields of study or from specific national or linguistic groups have no particular advantage. When topics have a national context, they refer to United States or Canadian history, culture, art, or literature. However, knowledge of these contexts is not needed to answer the structural or grammatical points in question.

  This section is divided into two parts. In the first part, which measures understanding of basic grammar, the examinee reads incomplete sentences printed in the test book. From the four responses provided for each incomplete sentence, the examinee must choose the word or phrase that best completes the given sentence. Only one of the choices fits correctly into the particular sentence. The second part tests knowledge of the grammar of written English. Here the examinee reads sentences in which some words or phrases are underlined. He or she must identify the one underlined word or phrase in each sentence that would not be accepted in standard written English.

- Section 3, Vocabulary and Reading Comprehension, tests the ability to understand the meanings and uses of words in written English as well as the ability to understand a variety of reading materials. So there is no advantage to individuals in any one field of study, the questions based on reading materials do not require outside knowledge of the subject matter.

  This section is divided into two parts. The vocabulary part contains sentences in which one word or phrase is underlined. Each sentence is followed by four choices consisting of single words or phrases, and the examinee must choose the one word or phrase that would best preserve the meaning of the original sentence if it were substituted for the underlined word(s). Although some of the choices may include words or phrases that would have the same meaning as the underlined portion if they were placed in different sentences, only one of the four choices will fit the sentence that is given.

  In the reading comprehension part of this section, the examinee must read a variety of short passages on academic subjects. Each passage is followed by questions about the meaning of the passage. Most questions are concerned with the main and secondary ideas in the passages, although questions based on inferences or analogies are included. In all cases, the questions can be answered by reading and understanding the passages.

Test of Written English

In 1986, the TOEFL program introduced the Test of Written English (TWE). This direct assessment of writing proficiency was developed in response to requests from many colleges, universities, and agencies that use TOEFL scores. The TWE test is currently (1990) a required section of the TOEFL test at four administrations a year. For more information about TWE, see page 30.

Development of TOEFL Test Questions

Material for the TOEFL test is written by language specialists, who are given rigorous training in writing questions for the test before they undertake actual writing assignments. Additional material is prepared by members of the TOEFL Committee of Examiners and by ETS test specialists. All item specifications, questions, and final test forms are reviewed internally at ETS for cultural and racial bias and content appropriateness, following established ETS procedures.

In addition, each final form of the TOEFL test is reviewed by at least one external consultant to ensure that the form is free of any language, symbols, or content that might be considered potentially offensive or inappropriate for major subgroups of the TOEFL test populations, or that might serve to perpetuate any negative attitude that could be conveyed to these subgroups. These consultants are specialists in cross-cultural psychology or anthropology and must have had significant experience living or working abroad (preferably in a non-Western culture, since most TOEFL examinees are from non-Western countries). All participate in a training workshop arranged by the TOEFL program before they serve as cultural sensitivity reviewers.

All questions are pretested on representative groups of foreign students. Only after the results of the pretested questions have been analyzed for statistical and content appropriateness are questions selected for the final test forms.

Following the administration of each new form of the test, a statistical analysis of the responses to each question is conducted. On rare occasions, when a question does not function as expected, it will be reviewed again by test specialists. After this review, the question may be deleted from the final scoring of the test. The statistical analyses also provide continuous monitoring of the level of difficulty of the test, the reliability of the entire test and of each section, intercorrelations among the sections, and the adequacy of the time allowed for each section. (See "Statistical Characteristics of the Test," page 22.)

The TOEFL Testing Programs

The TOEFL test is administered internationally through the official Friday and Saturday Testing Programs. It is also administered at local institutions around the world through the Institutional Testing Program. The latter program does not provide official TOEFL score reports, and scores are for use by the administering institution only.

Friday and Saturday Testing Programs

The official TOEFL test is given at centers around the world one day each month. Under the Friday Testing Program, formerly called the Special Center Testing Program, it is given on six Friday test dates during the testing year. Likewise, under the Saturday Testing Program, previously titled the International Testing Program, the test is given on scheduled Saturday dates six times a year.

The TOEFL office diligently attempts to make the test available to all individuals who require TOEFL scores. It will attempt to open a center for even one examinee if there is no center in the person's country or if an established center is too far away. In 1988-89, more than 1,300 centers located in 170 countries and areas were established for the Saturday (International) Testing Program to accommodate the more than 441,100 persons registered to take the test; 350 centers in more than 50 countries and areas were established for the more than 124,900 persons registered to take TOEFL under the Friday (Special Center) Testing Program.

Registration and administration procedures are identical for the Friday and Saturday programs. The test itself is also identical in terms of format and content. Score reports for administrations under both programs are comparable. More information about these testing programs can be found in the *Bulletin of Information for TOEFL/TSE*. (See page 34.)

Operation of the Friday and Saturday Testing Programs

The TOEFL program provides a total of twelve possible test dates—one each month. However, the actual number of administrations at any one center in a given country or area is scheduled according to demand and the availability of space and supervisory staff.

There are sometimes local scheduling conflicts with national or religious holidays. Although the TOEFL office attempts to avoid scheduling administrations of the test on these dates, it may be unavoidable in some cases.

Registration must be closed well in advance of each test date to ensure the delivery of test materials to the test centers. Registration deadline dates are about seven weeks before the test dates for centers outside the United States and Canada and five weeks before the test dates for centers within these two countries.

Almost all administrations are held as scheduled. On occasion, however, shipments of test materials may be impounded by customs officials or delayed by mail embargoes or transportation strikes. Other problems, ranging from political disturbances within countries, to power failures, to the last-minute illness of a test supervisor, may also force postponement of a TOEFL test administration.

If an administration must be canceled, a make-up administration is scheduled, usually on the next regularly scheduled test date. Occasionally it is necessary to arrange a make-up administration on another date.

A completely new form of the test is used at each administration. After an administration, the answer sheets are returned to ETS for scoring; test results are mailed to score recipients about one month after the answer sheets are received at ETS.

Institutional Testing Program (ITP)

The Institutional Testing Program is a flexible testing program that permits approved institutions throughout the world to administer TOEFL to their own students on dates convenient for them (except for regularly scheduled TOEFL administration dates), using their own facilities and staff. Each year a number of forms of the TOEFL test previously used in the Friday and Saturday Testing Programs are made available for the Institutional Program.

In addition to the regular TOEFL test, which is especially appropriate for use with students at the intermediate and higher levels of English language proficiency, the Institutional Program offers the Preliminary Test of English as a Foreign Language (Pre-TOEFL) for individuals at the beginning level. Pre-TOEFL measures the same components of English language skills as TOEFL. However, Pre-TOEFL is less difficult and shorter. Pre-TOEFL test results are based on a restricted scale that provides more discriminating measurements at the lower end of the TOEFL scale.

How Institutional TOEFL Can Be Used

The Institutional Testing Program is offered primarily to assist institutions in placing students in English courses at the appropriate level of difficulty, for determining whether additional work in English is necessary before an individual can undertake studies in that language, or as preparation for an official Friday or Saturday TOEFL administration. Conditions under which the Institutional Program test is given are governed by the administering institution, which must assume all responsibility for any irregularities.

Institutional TOEFL Test Scores

Scores earned under the Institutional Testing Program are comparable to scores earned under the worldwide Friday and Saturday testing programs. However, ITP scores are for use by the administering institution only. ETS reports test results to the administering institution in roster form, listing the names and scores (section and total) of all students who took the test at that administration. Two copies of the score record for each student are provided to the administering institution: a file copy for the institution and a personal copy for the individual. Both copies indicate that the scores were obtained at an Institutional Testing Program administration.

ETS does not report scores obtained under this program to other institutions as it does for official scores obtained under the Friday and Saturday Testing Programs. To ensure score validity, scores obtained under the Institutional Testing Program should not be accepted by other institutions to evaluate an individual's readiness to begin academic studies in English.

TOEFL Test Results

About one month after a TOEFL administration, test results are mailed to the examinees and to the official score recipients they have specified, provided that the answer sheets are received at ETS promptly after the administration.* Test results for examinees whose answer sheets are incomplete or whose answer sheets arrive late are usually sent two or three weeks later. All test results are mailed by the final deadline — twelve weeks after the test.

For the basic TOEFL test fee, each examinee is entitled to four copies of the test results: one copy is sent to the examinee, and up to three official score reports are sent directly by ETS to the institutions whose assigned code numbers the examinee marked on the answer sheet.†

The most common reason that institutions do not receive score reports following an administration is that examinees do not properly specify the institutions as score report recipients by marking the correct codes on the test answer sheet. (Examinees cannot write out the names of recipients on the answer sheet.) An examinee who does not mark the code number of an institution on the answer sheet must submit a Score Report Request Form naming the institution that is to receive the scores. There is a fee for this service.

A list of the most frequently used institution and agency codes is printed in the *Bulletin of Information*. An institution whose code number is not listed but that wishes to ensure that applicants submit official TOEFL score reports should give applicants its code number before they take the test. (This information should be included in application materials prepared for international students.)

Note: An institution that does not know its TOEFL code number or wishes to obtain one should write to Code Control, Educational Testing Service, Princeton, NJ 08541-0001, USA.

Examinee Score Records

Examinees receive their test results on a form titled Examinee's Score Record. These are NOT official TOEFL score reports and should not be accepted by receiving institutions.

Acceptance of Test Results Not Received from ETS

One should also bear in mind that score records received from any source other than ETS could have been altered. **In particular, photocopies of score records should not be accepted**. Institution and agency officials are urged to verify all TOEFL scores supplied by examinees by calling toll free 800-257-9547. (Officials in Alaska, Hawaii, Pennsylvania, and Canada should call collect to 215-750-8050.) TOEFL/TSE Services will either confirm or deny the accuracy of the scores submitted by examinees.

* Test results for the Institutional Testing Program are handled differently (see "Institutional Testing Program" above).

† An institution or agency that is sponsoring an examinee and has made prior arrangements with the TOEFL office will also receive a copy of the examinee's official score report if the examinee has given permission to the TOEFL office.

If there is a discrepancy between the official scores recorded at ETS and those submitted in any form by an examinee, the institution will be requested to send ETS a copy of the score record supplied by the examinee. At the written request of an official of the institution, ETS will report to it the official scores, as well as all previous scores recorded for the examinee within the last two years. (Also see "Test Score Data Retention" below.)

How to Recognize and Avoid Unofficial Score Reports

1. Never accept photocopies of score reports.
2. Always verify any personal copies sent by applicants.
3. Total score results always end in "0," "3," or "7."
4. The date of the appropriate *Test and Score Manual* is listed under interpretive information on the official score report.
5. For verification, photos of examinees are obtainable from the TOEFL program office.

Additional Score Reports

Individuals who have taken TOEFL at scheduled Friday or Saturday test administrations may request that official score reports be sent to additional institutions at any time up to two years after the date on which they took the test.

There are two score reporting services: (1) **regular** and (2) **rush** reporting. The regular service mails additional score reports within three weeks after receipt of an examinee's Score Report Request Form. The rush reporting service mails score reports to institutions within two working days after a request form has been received. There is an additional fee for the rush service.

Test Score Data Retention

Language proficiency can change considerably in a relatively short period. Therefore, the TOEFL office will not report scores that are more than two years old. Individually identifiable TOEFL scores are retained on score storage tapes for only two years from the date of the test. Individuals who took the TOEFL test more than two years ago must take it again if they want scores sent to an institution.* After two years, all information that could be used to identify an individual is removed from the database; other information is retained indefinitely. Score data that may be used at any time for research or statistical purposes do not include individual examinee identification information.

Confidentiality of Information

Information retained in TOEFL test files about an examinee's native country, native language, and the institutions to which the test scores have been sent, as well as the actual scores, is the same as the information printed on the examinee's score record and on the official score reports. An examinee's official score report will be sent only to those institutions or agencies designated on the answer sheet by the examinee on the day of the test, on a Score Report Request Form submitted at a later date, or otherwise specifically authorized by the examinee.†

If there is a discrepancy between the official scores recorded at ETS and those submitted to an institution by an examinee, ETS will report the official scores to that institution or agency. Examinees are advised of this policy in the *Bulletin,* and, in signing their completed registration forms, they accept these conditions.

Score users bear the responsibility of maintaining confidentiality of an individual's score information.

Contents of TOEFL Score Reports

TOEFL score reports give the score for each of the three sections of the test and the total score. Examinees who take the TOEFL test during an administration at which the Test of Written English is given will also receive a TWE score printed in a separate field on the TOEFL score report. Facsimiles of the official score report and the Examinee's Score Record are reproduced on pages 11 and 12.

Fields of Graduate Study Other Than Business or Law

HUMANITIES
11 Archaeology
12 Architecture
26 Art History
13 Classical Languages
28 Comparative Literature
53 Dramatic Arts
14 English
29 Far Eastern Languages and Literature
15 Fine Arts, Art, Design
16 French
17 German
04 Linguistics
19 Music
57 Near Eastern Languages and Literature
20 Philosophy
21 Religious Studies or Religion
22 Russian/Slavic Studies
23 Spanish
24 Speech
10 Other foreign languages
98 Other humanities

SOCIAL SCIENCES
27 American Studies
81 Anthropology
82 Business and Commerce
83 Communications
84 Economics
85 Education (including M.A. in Teaching)
01 Educational Administration
70 Geography
92 Government
86 History
87 Industrial Relations and Personnel
88 International Relations
18 Journalism
90 Library Science
91 Physical Education
97 Planning (City, Community, Regional, Urban)
92 Political Science
93 Psychology, Clinical
09 Psychology, Educational
58 Psychology, Experimental/ Developmental
79 Psychology, Social
08 Psychology, other
94 Public Administration
95 Social Work
96 Sociology
80 Other social services

BIOLOGICAL SCIENCES
31 Agriculture
32 Anatomy
05 Audiology
33 Bacteriology
34 Biochemistry
35 Biology
45 Biomedical Sciences
36 Biophysics
37 Botany
38 Dentistry
39 Entomology
46 Environmental Science
40 Forestry
06 Genetics
41 Home Economics
25 Hospital and Health Services Administration
42 Medicine
07 Microbiology
74 Molecular and Cellular Biology
43 Nursing
77 Nutrition
44 Occupational Therapy
56 Pathology
47 Pharmacy
48 Physical Therapy
49 Physiology
50 Public Health
55 Speech-Language Pathology
51 Veterinary Medicine
52 Zoology
30 Other biological sciences

PHYSICAL SCIENCES
54 Applied Mathematics
61 Astronomy
62 Chemistry
78 Computer Sciences
63 Engineering, Aeronautical
64 Engineering, Chemical
65 Engineering, Civil
66 Engineering, Electrical
67 Engineering, Industrial
68 Engineering, Mechanical
69 Engineering, other
71 Geology
72 Mathematics
73 Metallurgy
75 Oceanography
76 Physics
59 Statistics
60 Other physical sciences

Use 99 for any department not listed.

* A TOEFL score is measurement information and is subject to all the restrictions noted in this *Manual.* (These restrictions are also noted in the *Bulletin of Information.*) The score is not the properly of the examinee.

† See second footnote on page 9.

In addition to test scores, native country, native language, and birth date, the report includes other personal information about the examinee and information about the test. The institution code designates the recipient college, university, or agency, and the department code number ("Dept. Code") identifies the professional school, division, department, or field of study in which the graduate applicant plans to enroll. The department code list shown on page 10 is also included in the *Bulletin of Information*. The department code for all business schools is (02), for law schools (03), and for "unlisted departments" (99).

TOEFL Scores

The raw scores for the three sections of the TOEFL test are the number of questions answered correctly. No penalty points are subtracted for wrong answers. Although each new form of the test is constructed to match previous forms, the level of difficulty may vary slightly from form to form. Since the number of questions answered correctly depends in part on the difficulty level of the test, the raw scores on each section are converted to a uniform scale by a method that takes into account the level of difficulty of the particular test form.

At the time of the first administration of the three-section TOEFL test (1976), the scale for reporting the total score was linked to the scale that was in use for the original five-section test. That scale had been established using score data for a group of foreign students tested in February 1964.*

For the current three-section test, the three separate sections are scaled so the mean scaled score for each section equals one tenth of the total scaled score mean (the standard deviations of the scaled scores for the three sections are equal) and the total score equals ten-thirds times the sum of the three section scaled scores.

* The number-right scores for the five sections had been converted to a scale such that the distribution of converted scores had a mean of 50 and a standard deviation of 10. The total score was twice the sum of the five section converted scores and thus had a mean of 500 for the February 1964 group. Total scores for both the five-section and three-section tests have fallen within the range of 200 to 700.

TOEFL section scores are reported on a scale that can range from 20 to 68. TOEFL total scores are reported on a scale that can range from 200-677.

	Section 1		Section 2		Section 3		Sum
Example:	46	+	54	+	50	=	150
			(150 × 10) ÷ 3 = 500				

This method of scaling results in rounded scores for which the last digit can take on only three values: zero, three, or seven.

Scores for each new test form are converted to the same scale by a statistical procedure known as score equating that determines equivalent scaled scores for persons of equal ability regardless of the level of difficulty of the form of the test and the average level of ability of the group taking the test. Prior to September 1978, score equating was done by including in each new form a set of items from a previous form. In order to improve test security and thus the accuracy of the equating process, a feasibility study of item response theory equating was undertaken in July 1977 (Cowell, 1982). Based on the success of this study, the item response theory method has been used to equate TOEFL test forms since September 1978.†

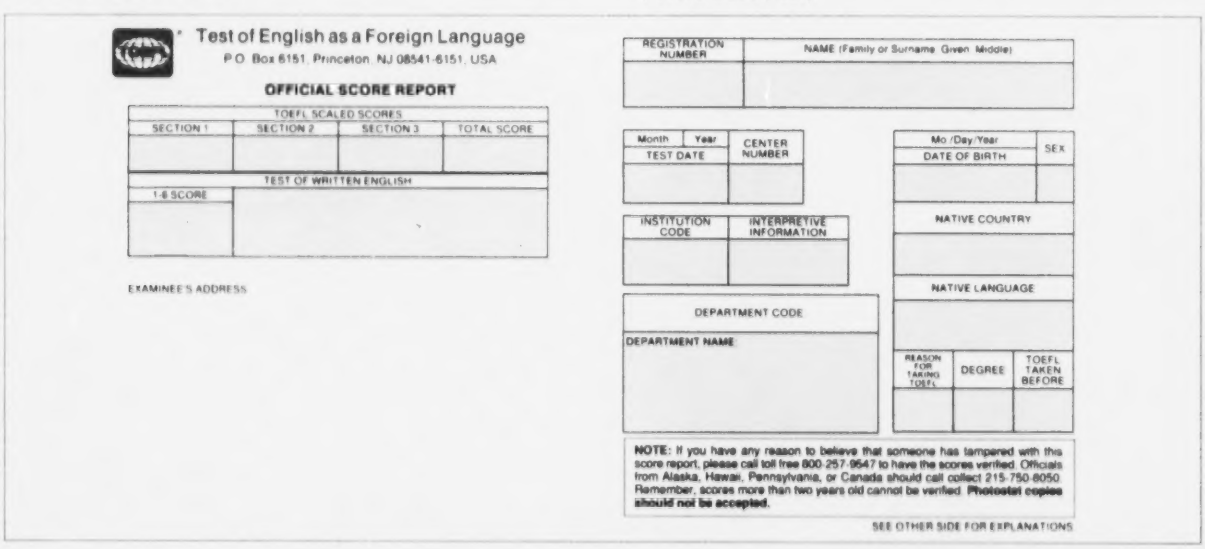
The reported scores are neither the number nor the percentage of questions answered correctly. They are not related to the distribution of scores on any other test, such as the SAT or the GRE tests.

Actual ranges of observed scores for the period from July 1987 through June 1989 are shown in Table 1.

Table 1. Minimum and Maximum Observed Section and Total Scores, July 1987-June 1989

Section	*Min.*	*Max.*
1. Listening Comprehension	23	68
2. Structure and Written Expression	20	68
3. Vocabulary and Reading Comprehension	20	67
Total Score	210	677

† Further information about item response theory is given in works by McKinley (1989) and Lord (1980).



Test of English as a Foreign Language
P.O. Box 6151, Princeton, NJ 08541-6151, USA

OFFICIAL SCORE REPORT

TOEFL SCALED SCORES			
SECTION 1	SECTION 2	SECTION 3	TOTAL SCORE

TEST OF WRITTEN ENGLISH	
1-6 SCORE	

EXAMINEE'S ADDRESS

REGISTRATION NUMBER	NAME (Family or Surname, Given, Middle)

Month / Year TEST DATE	CENTER NUMBER

Mo./Day/Year DATE OF BIRTH	SEX

INSTITUTION CODE	INTERPRETIVE INFORMATION

NATIVE COUNTRY

NATIVE LANGUAGE

DEPARTMENT CODE

DEPARTMENT NAME

REASON FOR TAKING TOEFL	DEGREE	TOEFL TAKEN BEFORE

NOTE: If you have any reason to believe that someone has tampered with this score report, please call toll free 800-257-9547 to have the scores verified. Officials from Alaska, Hawaii, Pennsylvania, or Canada should call collect 215-750-8050. Remember, scores more than two years old cannot be verified. **Photostat copies should not be accepted.**

SEE OTHER SIDE FOR EXPLANATIONS

INFORMATION ABOUT SCORES: Three section scores and a total score are reported for the current form of the TOEFL test. The three sections are:

Section 1—Listening Comprehension
Section 2—Structure and Written Expression
Section 3—Vocabulary and Reading Comprehension

TEST OF WRITTEN ENGLISH (TWE): September, October, March, and May administrations.

Scores	Explanations of TWE Scores
6	Clearly demonstrates competence in writing on both the rhetorical and syntactic levels, though the essay may have occasional errors.
5	Demonstrates competence in writing on both the rhetorical and syntactic levels, though the essay will have occasional errors.
4	Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.
3	Demonstrates some developing competence in writing, but the essay remains flawed on either the rhetorical or syntactic level, or both.
2	Suggests incompetence in writing.
1	Demonstrates incompetence in writing.
INR	Did not write an essay.
OFF	Did not write on assigned topic.

REASON FOR TAKING TOEFL

1 = To enter a college or university as an undergraduate student
2 = To enter a college or university as a graduate student
3 = To enter a school other than a college or university
4 = To become licensed to practice a profession
5 = To demonstrate proficiency in English to the company for which the examinee works or expects to work
6 = Other than above
0 = Not answered

PLANS TO WORK FOR DEGREE

1 = Yes 2 = No 3 = Not answered

NUMBER OF TIMES TOEFL TAKEN BEFORE

1 = One 2 = Two 3 = Three 4 = Four or more 0 = None or not answered

INTERPRETIVE INFORMATION: The copyright date of the current edition of the *TOEFL Test and Score Manual* is printed here. Individuals responsible for evaluating TOEFL scores should refer to the *Manual*, which is available on request from the TOEFL Program Office, P.O. Box 6155, Princeton, NJ 08541-6155, USA.

TEST DATE: Because English proficiency can change considerably in a relatively short period, please note the date on which the test was taken. **Scores more than two years old cannot be reported, nor can they be verified.**

To ensure that a score report has not been tampered with or altered, check the following:

Is a different type face used in any area?

Is the last digit of the total score 0, 3, or 7? (See "Reading TOEFL Score Reports.")

Are there erasures? Do any of the shaded areas seem lighter than others, or are any of these areas blurred?

Test of English as a Foreign Language
P.O. Box 6151, Princeton, NJ 08541-6151, USA

EXAMINEE'S SCORE RECORD

TOEFL SCALED SCORES			
SECTION 1	SECTION 2	SECTION 3	TOTAL SCORE
TEST OF WRITTEN ENGLISH			
1-6 SCORE			

EXAMINEE'S ADDRESS

REGISTRATION NUMBER	NAME (Family or Surname, Given, Middle)

Month	Year	CENTER NUMBER	SPONSOR CODE
TEST DATE			

Mo./Day/Year	SEX
DATE OF BIRTH	

INSTITUTION CODE	DEPT. CODE

NATIVE COUNTRY

NATIVE LANGUAGE

REASON FOR TAKING TOEFL	DEGREE	TOEFL TAKEN BEFORE

SEE OTHER SIDE FOR EXPLANATIONS

NOTE FOR INSTITUTIONS: This score record has been supplied by the examinee. Score recipients are urged to verify this score record by calling toll free 800-257-9547. From Alaska, Hawaii, Pennsylvania, or Canada call collect 215-750-8050. Scores more than two years old cannot be verified. **Photostat copies should not be accepted.**

TEST DATE: This is the date on which you took the TOEFL test. Because English proficiency can change considerably in a relatively short period, the TOEFL office cannot report scores that are more than two years old.

INFORMATION ABOUT SCORES: Three section scores and a total score are reported for the current form of TOEFL. The three sections are:

Section 1—Listening Comprehension
Section 2—Structure and Written Expression
Section 3—Vocabulary and Reading Comprehension

Test of Written English (TWE): September, October, March, and May administrations.

NOTE FOR THE EXAMINEE: The institution code numbers printed on your score record are the ones you marked correctly on your answer sheet. If you wrote any numbers on your answer sheet that are not shown on this score record, the numbers you gave were not correct and the TOEFL office was not able to send those score reports. To have official score reports sent to those institutions or agencies, follow the directions under "Score Reporting Services" in the *Bulletin*.

If code number 1000 is printed on this score record, your official score report was sent to a recipient that has no code number but whose name and address you printed on a Score Report Request Form.

Also, if you did not send enough money with your Score Report Request Form, the TOEFL office sent reports only to as many institutions as you paid for, **no matter how many you requested.**

SCORES MORE THAN TWO YEARS OLD CANNOT BE REPORTED.

REASON FOR TAKING TOEFL

1 = To enter a college or university as an undergraduate student
2 = To enter a college or university as a graduate student
3 = To enter a school other than a college or university
4 = To become licensed to practice a profession
5 = To demonstrate proficiency in English to the company for which you work or expect to work
6 = Other than above
0 = Not answered

PLANS TO WORK FOR DEGREE
1 = Yes
2 = No
3 = Not answered

NUMBER OF TIMES TOEFL TAKEN BEFORE
1 = One
2 = Two
3 = Three
4 = Four or more
0 = None or not answered

DEPT.	WHERE REPORT WAS SENT
02	Report was sent to the admissions office of a graduate school of management (business).
03	Report was sent to the admissions office of a graduate school of law.
04-99	Report was sent to the admissions office for graduate study in a field other than management (business) or law according to the codes you marked on your answer sheet on the day of the test. See the *Bulletin* for these codes.
00	Report was sent to the admissions office for undergraduate study (00 is also printed after the institution code number for each institution or agency that is not a college or university).

TWE Scores

The Test of Written English (TWE) is scored holistically using a six point, criterion-referenced scoring guide. Following each TWE administration, the essays are scored by trained readers who are English and ESL writing specialists affiliated with accredited U.S. and Canadian colleges, universities, and secondary schools. All readers have demonstrated their ability to read reliably using the TWE Scoring Guide.

All TWE readings are conducted in a central location to ensure that standardized reading procedures are maintained. The essays are randomized before each round of scoring to minimize context effect (the tendency of a reader to compare an essay with those preceding it). The readers read and score the essays using sample papers scored by the reading managers to anchor the essays to the scoring guide. A score of 6 demonstrates clear competence in writing, while a score of 1 demonstrates incompetence in writing or an unattempted essay (the score report distinguishes between these two possibilities with a special code).

Each essay is read by two readers. If both agree on the score assigned (for example, Reader 1 score = 3, Reader 2 score = 3), that score is reported. If the scores assigned are within one point of each other (Reader 1 score = 3, Reader 2 score = 4), the score assigned is the average of the two scores (in this case, a score of 3.5). If the scores differ by more than one point (for example, Reader 1 score = 3, Reader 2 score = 5), the essay is adjudicated by the reading managers, who are the most experienced readers. Essays that are not written to the topic are not scored and are reported as off-topic responses from which no estimate of writing proficiency could be established.

Carefully controlled topic development, pretesting, pretest analysis, and postadministration scoring procedures maintain the comparability of TWE scores within and across administrations. For information about the reliability of TWE scores, see page 25.

Hand-Scoring Service

Examinees are responsible for properly completing their answer sheets to ensure accurate scoring. They are instructed to use a medium-soft black lead pencil, to mark only one answer to each question, to fill in the answer space completely so the letter inside the space cannot be seen, and to erase all extra marks thoroughly. Failure to follow any of these instructions may result in the reporting of an inaccurate score.

Examinees who question whether their reported scores are accurate may request that their answer sheets be hand scored. There is a fee for this service. A request for hand scoring must be received within six months of the test date; later requests cannot be honored.

The TOEFL office has established the following hand-scoring procedures: the answer sheet to be hand-scored is first confirmed as being the one completed by the person requesting this service; the answer sheet is then scored twice by hand, by specially trained ETS staff working independently. If there is a discrepancy between the hand-scored and computer-scored results, the hand-scored results — *which may be higher or lower than those originally reported* — will be reported to all recipients of the earlier scores, and the hand-scoring fee will be refunded to the examinee. The results of the hand scoring are available about three weeks after receipt of the examinee's request. In general very few score changes have resulted from hand-scoring requests.

Scores of Questionable Validity

Improved scores over time are to be expected if a person is studying English and may not indicate irregularities. However, institutions that note inconsistencies between test scores and English performance may request copies of the examinees' TOEFL photo file records to verify that they are the persons who took the test (see "Examinee Identification Service," page 35). Institutions should notify the TOEFL office if they

find any evidence of impersonation.

Irregularities uncovered by institutions and reported to ETS, as well as those brought to the attention of the TOEFL office by examinees or supervisors who believe that misconduct may have taken place, are investigated. In addition, test results are routinely monitored.

Misconduct irregularities are reviewed, and scores may be canceled by ETS. For other irregularities, the ETS Test Security Office assembles relevant documents, such as previous score reports, registration forms, and answer sheets. When handwriting differences or evidence of possible copying or exchange of answer sheets is found, the case is referred to the ETS Board of Review, a group of senior professional staff members. Based on its independent examination of the evidence, the Board of Review directs appropriate action.

ETS policy and procedures are designed to provide reasonable assurance of fairness to examinees in both the identification of suspect scores and the weighing of information leading to possible score cancellation. These procedures are intended to protect both score users and examinees from inequities that could result from decisions based on fraudulent scores and to maintain the integrity of the test.

Use of TOEFL Test Scores

The TOEFL test is a measure of English proficiency. It is not a test of academic aptitude or of subject matter competence. TOEFL test scores can assist in determining whether an applicant has attained sufficient proficiency in English to study at an institution. However, even a student with native or near-native English language ability who is not academically prepared for a program of study may not easily succeed even though he or she may achieve a high TOEFL score. Therefore, determination of academic admissibility of nonnative English speakers is dependent upon numerous other factors, such as previous academic record, other institution(s) attended, level and field of study, motivation, and cultural background.

If a nonnative English speaker meets academic requirements, official TOEFL test scores may be used in making the following kinds of decisions:

1. The applicant may begin academic work with no restrictions.
2. The applicant may begin academic work with some restrictions on academic load and in combination with concurrent work in English. (This decision implies that the institution can provide the appropriate English courses to complement the applicant's part-time academic schedule.)
3. The applicant is declared eligible to begin an academic program within a stipulated period of time but is assigned to a full-time English program. (Normally, such a decision is made when an institution has its own intensive English program.)
4. The applicant's official status will not be determined until he or she reaches a satisfactory level of English proficiency. (Such a decision will require that the applicant pursue full-time English training, at the same institution or elsewhere.)

All of the above decisions require the institution to judge whether the applicant has sufficient command of English to meet the demands of a regular or modified program of study. Such decisions should never be based on TOEFL test scores alone; they should be based on all relevant information available.

Who Should Take TOEFL?

All nonnative speakers of English should provide evidence of their level of English proficiency prior to beginning academic work at an institution where English is the language of instruction. Most of the institutions that responded to a survey conducted by TOEFL in 1989 reported requiring TOEFL scores for the following categories of applicants.

- Individuals from countries in which English is one of the official languages, but not necessarily the first language of the majority of the population or the language of instruction at all levels of schooling. Such countries may include, but are not limited to, the British Commonwealth countries and U.S. territories and possessions.

- Persons from countries where English is not the native language, even though there may be schools or universities in which English is the language of instruction.

Many institutions reported that they frequently do not require TOEFL test scores of certain kinds of foreign applicants. These include:

— Nonnative speakers who hold degrees or diplomas from postsecondary institutions in English-speaking countries (e.g., the United States, Canada, England, the Republic of Ireland, Australia, and New Zealand), provided they have spent a specific minimum period of time in successful full-time study (generally two years) with English as the medium of instruction.

— Transfer students from other institutions in the United States or Canada after favorable evaluation of previous academic course work and course load and length of time at the previous institution.

— Nonnative speakers who have taken TOEFL within the past two years and who have successfully pursued academic work in an English-speaking country for a specified minimum period of time (generally two years) with English as the medium of instruction.

Guidelines for Using TOEFL Test Scores

As part of its general responsibility for the tests it produces, the TOEFL program is concerned about the use of TOEFL test scores by recipient institutions. The program office makes every effort to ensure that institutions use TOEFL test scores properly—for example, by providing this *Manual* to all institutions that are interested in using the scores and by regularly advising institutions of any program changes that may affect the interpretation of TOEFL test scores. The TOEFL office encourages individual institutions to request its assistance relating to the proper use of scores.

An institution that uses TOEFL test scores should consider certain factors in evaluating an individual's performance on the test and in determining appropriate score requirements. The following guidelines are presented to assist institutions in arriving at reasonable decisions.

Base the evaluation of an applicant's readiness to begin academic work on all available relevant information, not solely on TOEFL test scores.

The TOEFL test measures an individual's ability in several areas of English language proficiency. It is not designed to provide information about scholastic aptitude, motivation, language-learning aptitude, or cultural adaptability. The eligibility of a foreign applicant should be fully established on the basis of all relevant academic and other criteria, including sufficient proficiency in English to undertake the academic program at the institution in question.

Do not use rigid "cut-off" scores in evaluating an applicant's performance on the TOEFL test.

Because test scores are not perfect measures of ability, the use of rigid cut-off scores should be avoided. The standard error of measurement should be understood and taken into consideration in making decisions about an individual's test performance or in establishing appropriate critical score ranges for the institution's academic demands (see "Reliabilities and the Standard Error of Measurement," page 25).

Consider section scores as well as total scores.

The total score on the TOEFL test is based on the scores on the three multiple-choice sections of the test. While a number of applicants may achieve the same total score, they may have different section scores, which could be significant. For example, an applicant with a low score on the Listening Comprehension section but relatively high scores on the other sections might have greater initial difficulty in lecture courses.* This information could be used in the advising and placement of applicants.

If an applicant's score on the Structure and Written Expression section is considerably lower than the scores on the other sections or if the applicant's score on the TWE is low, it may be that the individual should take a reduced academic load or be placed in a course designed to improve composition skills and knowledge of English grammar. An applicant whose score on the Vocabulary and Reading Comprehension section is much lower than the scores on the other two sections might be advised to take a reduced academic load or to postpone enrollment in courses that involve a significant amount of reading.

Consider the kinds and levels of English proficiency required in different fields and levels of studies and the resources available at the institution for improving the English language skills of nonnative speakers.

An applicant's field of study can affect the kind and level of language proficiency that is appropriate. Students pursuing studies in fields requiring high verbal ability (such as journalism) will need a greater command of English, particularly structure and written expression and writing, than will those in fields that are not so dependent upon reading and writing abilities. Many institutions require a higher range of TOEFL test scores for graduate applicants than for undergraduates (see "Survey of Institutional Policies Regarding TOEFL," page 17).

Institutions offering courses in English for foreign students can modify academic course loads to allow for additional concurrent language training, and thus may be able to consider applicants with a lower range of scores than can institutions that do not offer additional language training.

Consider TOEFL test scores in interpreting an applicant's performance on other standardized tests.

Foreign applicants are frequently required to take standardized admission tests in addition to the TOEFL test. In such instances, TOEFL scores may prove useful in interpreting the scores obtained on the other tests. For example, if an applicant's TOEFL scores are low and the scores on another test are also low (particularly one that is primarily a measure of aptitude or achievement in verbal areas), one can legitimately infer that the applicant's performance on the other test was impaired because of deficiencies in English. On the other hand, applicants with high verbal aptitude scores but low TOEFL scores should be reviewed carefully. The scores may be wrong or false.

Interpreting the relationship between the TOEFL test and aptitude and achievement tests in verbal areas can be complex. Few of even the most qualified foreign applicants approach native proficiency in English. Factors such as cultural and international differences in educational programs may also affect performance on tests of verbal ability.

The TOEFL program has published four research reports that can assist in evaluating the effect of language proficiency on an applicant's performance on specific standardized tests. *The Performance of Nonnative Speakers of English on TOEFL and Verbal Aptitude Tests* (Angelis, Swinton, and Cowell, 1979) gives comparative data about foreign student performance on TOEFL and either the GRE verbal or the SAT verbal and the Test of Standard Written English (TSWE). It provides interpretive information about how combined test results might best be evaluated by institutions that are considering foreign students. *The Relationship between Scores on the Graduate Management Admission Test* and the *Test of English as a Foreign Language* (Powers, 1980) provides a similar comparison of performance on the GMAT and TOEFL. Finally, *Language Proficiency as a Moderator Variable in Testing Academic Aptitude* (Alderman, 1981) and *GMAT and GRE*

* See page 31 for information about the Test of Spoken English and oral proficiency.

Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL (Wilson, 1982) contain information supplementing that provided in the other two studies. (See "Validity," page 26.)

Do not use TOEFL test scores for predicting academic performance.

The TOEFL test is a measure of English language proficiency, not of academic aptitude. It is not appropriate to use TOEFL scores to predict academic performance. Numerous predictive validity studies,* using grade-point averages as criteria, have been conducted in the past. These studies have shown that correlations between TOEFL test scores and grade-point averages are too low to be of any practical significance. Moreover, low correlations are to be expected when TOEFL scores are used properly. If an institution admits to academic work only those foreign applicants who have demonstrated a high level of language competence, one would expect that English proficiency would not be a major factor in influencing subsequent academic success.

The English proficiency of a foreign applicant is not as stable a characteristic as verbal or mathematical aptitude. Proficiency in a language is subject to change over relatively short periods of time. If considerable time has passed between the date on which an applicant took the TOEFL test and the date on which he or she actually begins academic studies, there may be a greater impact on academic performance due to language loss than had been anticipated. On the other hand, a student who might be handicapped because of language problems during the first term of study might not be handicapped in subsequent terms.

Assemble information on the validity of TOEFL test score requirements at the institution.

The establishment of appropriate standards of language proficiency through the use of TOEFL scores can have a favorable effect on the success rate of the foreign student population at an institution. However, such standards should be supported by the collection of data based upon the student population at the particular institution. This information may be useful in raising or lowering the standard as necessary. It may also be used to defend the standard should its legitimacy be challenged.

Institutions that use TOEFL scores should collect information about subsequent performance by applicants who are admitted. Student scores may be compared to a variety of criterion measures, such as students' ratings of the adequacy of their language skills for formal study in English, or instructors' ratings of the adequacy of students' language skills. When evaluating a standard with data obtained solely from students who have met the standard (that is, only students who have been admitted), one encounters an interesting phenomenon. If the current standard is set at a high level, so only those with a high degree of language proficiency are admitted, there may be no relationship between the TOEFL scores and any of the criterion measures. Because there will be no important variability in English proficiency among the group members, variations in success on the criterion variable will be due to other causes, such as knowledge of the subject matter, academic aptitude, study skills, cultural adaptability, and financial security.

On the other hand, if the standard is set at a low level, a large number of applicants selected with TOEFL scores may be unsuccessful because of inadequate command of English, and the scores will correlate quite strongly with the criterion. Thus, with a standard that is neither high nor low, the correlation between TOEFL scores and subsequent success will be moderately low. The magnitude of the correlation will depend on other factors as well. These factors may include variability in scores on the criterion measure, and/or the reliability of the raters, if raters are used. Expectancy tables can be used to show the distribution of performance on the criterion variables for students with given TOEFL scores. Thus, it may be possible to depict the number or percentage of students at each score level who attain a certain language proficiency rating as assigned by an instructor, or who rate themselves as not being hampered by lack of English skills while pursuing college-level studies.

Another approach is to use a regression equation to support a score standard. In such a case, one may simply solve the equation generated during the correlational analysis in order to determine the TOEFL score needed so the average language proficiency rating assigned by the instructor will be at a certain level. Additional information about the setting and validation of test score standards is available in a report by Livingston and Zieky (1982).

Several other methodological issues should be considered when conducting a standard setting or validation study. Since language proficiency can change within a relatively short time, student performance on a criterion variable should be assessed during the first term of enrollment. The fact that an individual exhibits deficient language skills during one term does not mean that the same deficiency will be apparent in subsequent terms. Also, if TOEFL scores are not obtained immediately prior to admission, gains or losses in language skills may reduce the relationship between the TOEFL test and the criterion.

Another issue that should be addressed is the relationship between subject matter or level of study and language proficiency. All subjects may not require the same level of language proficiency in order for the student to perform acceptably. For instance, the study of mathematics may require a lesser degree of English language proficiency than the study of philosophy. Similarly, freshman undergraduates who are required to take courses in a wide range of subjects may require a level of language proficiency different from that of graduate students who are concentrating on a specific field.

Section scores may also be taken into consideration in the setting and validating of score standards. For fields that require much reading, the Vocabulary and Reading Comprehension score may be particularly important. In fields in which little writing is required, the Structure and Written Expression or TWE score may be less important. Assessment of the relationship of subscores to the criterion variables can further refine the process of interpreting TOEFL scores.

To be useful, data about subsequent performance must be collected for relatively large numbers of students over an

* Chase and Stallings, 1966; Hamburg, 1979; Heil and Aleamoni, 1974; Hwang and Dizney, 1970; Odunze, 1982; Schrader and Pitcher, 1970; Sharon, 1972.

extended period of time. Institutions that have only a small number of foreign applicants each year or that have only recently begun to require TOEFL scores may not find it feasible to conduct the recommended studies. Such institutions might find it helpful to seek information and advice from colleges and universities that have had more extensive experience with the TOEFL test. The TOEFL office suggests that institutions evaluate their TOEFL requirements regularly to ensure that they are consistent with the institutions' own academic requirements and the language training resources they can provide nonnative speakers of English.

Examinees with Disabilities

Examinees with visual or physical impairments may use readers and/or amanuenses and may take the test in a separate room. Extended time and rest breaks are also permitted in nonstandard administrations. Examinees with hearing disabilities may take only Sections 2 and 3 of the TOEFL test and the TWE test; scores are reported only for those sections. A hearing-impaired examinee may take Sections 2 and 3 with the other examinees, or the test may be given in a separate room with extended time and rest breaks.

Nonstandard administrations are given on regularly scheduled test dates, and security procedures are the same as those followed for standard administrations.

The TOEFL office advises institutions that the test may not provide a valid measure of the examinee's proficiency, even though the conditions were designed to minimize any adverse effects of the examinee's disability upon test performance. The TOEFL office continues to recommend that alternative methods of evaluating English proficiency be used for individuals who cannot take the test under standard conditions. Criteria such as past academic record (especially if English has been the language of instruction), recommendations from language teachers or others familiar with the applicant's English proficiency, and/or a personal interview or evaluation are suggested in lieu of TOEFL scores. Because the individual circumstances of nonstandard administrations vary so widely and the number of examinees testing under nonstandard conditions is still quite small, the TOEFL program is not able to provide specific normative data for interpreting scores earned in such administrations.

A statement that the scores were obtained under nonstandard conditions is printed on the official score report (and on the Examinee's Score Record) of an examinee for whom special arrangements were made. Each score recipient is also sent an explanatory notice emphasizing that there are no normative data for scores obtained under nonstandard testing conditions and, therefore, that such scores should be used cautiously.

Appropriate Use of TOEFL Scores

In its efforts to ensure appropriate use of TOEFL scores, the TOEFL program maintains continuing contact with institutions regarding their use of the test. Formal contact is maintained through the biannual survey of institutional score users, conducted as part of the preparation of each new edition of the *Manual*. The TOEFL staff maintains informal contact with score users through professional meetings, conferences, and correspondence.

Survey of Institutional Policies Regarding TOEFL

The most recent survey of institutional score users was conducted in the fall of 1989. Questionnaires were sent to more than 2,200 institutions that use TOEFL scores. Survey recipients included two-year and four-year schools as well as graduate and professional schools and organizations. Completed questionnaires were returned by more than 800 institutions, all from the United States and Canada. Results of the statistical analysis of the 761 usable questionnaires are summarized below.

Survey Results

— How is the TOEFL test used in the admissions process?

The survey questionnaire asked institutions to indicate the purposes for which TOEFL scores are used. Of all respondents, 99 percent (750 of 761) indicated that TOEFL scores are used to determine sufficient English language proficiency to begin academic work. TOEFL was used for placement in or exit from an English language program by 19 percent of those responding, and 17 percent indicated that TOEFL was used to judge the validity of an applicant's scores on other standardized tests used in the admissions process (e.g., GRE, GMAT, SAT). These uses are consistent with established guidelines for using TOEFL scores (see page 15).

— Who is required to take the TOEFL test?

Of 386 undergraduate institutions responding, 346 (90 percent) indicated that all undergraduates who are not native speakers of English must provide TOEFL scores as evidence of their level of English proficiency; 24 schools required TOEFL scores only for some departments. Of 375 graduate schools, 363 (97 percent) said that TOEFL scores were required of all nonnative English-speaking undergraduate applicants.

— What types of students might be exempted from taking TOEFL?

The survey asked admissions officers to identify the types of students for whom TOEFL scores may not be required. Graduate and undergraduate institutions provided the following rank ordering of types of students whom the institutions may exempt from submitting TOEFL scores:

(564) (1) Those holding degrees or diplomas from institutions in English-speaking countries.

(300) (2) Those transferring from institutions in English-speaking countries.

(162) (3) Those who have successfully completed at least two years of college-level work in an English-speaking country with English as the language of instruction.

(96) (4) Those from countries in which English is an official language but not necessarily the language of instruction.

(81) (5) Those who have completed ESL instruction at an English language institute.

— How does an institution determine what TOEFL score ranges it should require or recommend of applicants?

Score users are advised to conduct in-house validity studies to determine the level of English language proficiency needed to successfully complete an academic program at that institution. Considered in the validity study are such criteria as the institution's past history with students who have similar test score ranges and the institution's resources for providing supplemental language education to applicants who, aside from their lack of English skills, meet the institution's academic requirements.

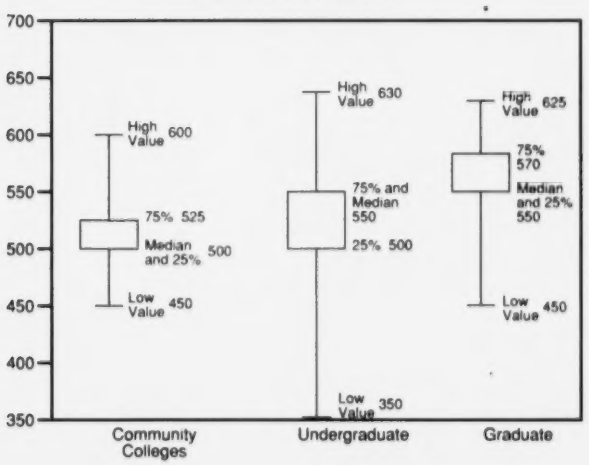
Of 729 institutions that responded to the question "Are you currently conducting or have you already conducted an in-house validity study of the TOEFL?" only 26 (4 percent) indicated that they had completed validity studies, and 53 institutions (7 percent) reported that they were planning or conducting in-house validity studies of their TOEFL policies. Of 744 respondents, 410 (55 percent) requested that the TOEFL program provide them with informational materials about conducting a local validity study. (For more information about local validity studies, see "Guidelines for Using TOEFL Test Scores," p.15).

— If TOEFL scores are required of foreign applicants, is there a minimum acceptable total score range?

Each institution that uses TOEFL scores sets its own policy about which score ranges will be required for each of its courses of study. The TOEFL program does not establish cut scores, nor does it determine which students must take the test.

As in the 1986 survey, community colleges surveyed in 1989 reported that, when considering a foreign student for admission, the minimum total TOEFL score they most frequently associate with the ability to begin academic work without restriction is 500. Undergraduate institutions reported that the minimum TOEFL total score they most frequently associate with this category is 550. For graduate and professional institutions, this ability level is still most frequently associated with a score of 550.

Box Plots of Minimum TOEFL Score Required for Beginning Academic Work with No Restrictions



As well as asking institutions to indicate the TOEFL total score range required to begin full-time academic undergraduate or graduate work with no restrictions, the survey asked what TOEFL scores were required for students to begin academic work in selected fields not requiring high verbal ability, limited academic work with supplemental English instruction, and full-time English language instruction prior to academic enrollment. The chart on page 19 summarizes the reported institutional policies.

— Do institutions ever change their TOEFL score requirements?

The questionnaire also asked if the institution had recently reevaluated its policy concerning a specific minimum total TOEFL score required of foreign students applying to the academic program. Of those respondents indicating a change in their TOEFL score requirements over the last three years, 34 undergraduate institutions and 60 graduate institutions had raised the minimum acceptable TOEFL score. Only 5 undergraduate institutions and 4 graduate institutions had lowered their TOEFL requirements. This shows a general tightening of TOEFL score requirements between 1986 and 1989 similar to that reported in the 1982, 1984, and 1986 surveys of TOEFL score users. The 1989 survey, then, is the fourth survey within nine years to show an increase in the level of TOEFL total score requirements.

— Do two-year schools use TOEFL scores differently than four-year schools?

Results from the 1982, 1984, and 1986 surveys showed that about 67 percent of responding community colleges required a total TOEFL score between 500 and 550 for academic admission. In the 1989 survey, 64 percent of the 28 community colleges responding indicated that they required a total TOEFL score between 500 and 550 for academic admission.

As in the 1982, 1984, and 1986 surveys, the 1989 survey respondents reported higher requirements for graduate applicants and for applicants in fields requiring high verbal ability, such as English literature, history, and business management.

— How are the TOEFL section scores used?

The questionnaire also asked institutions to indicate which section score, if any, is most significant in admissions decisions. Only 16 percent (117 of 743 respondents) indicated that emphasis was given to any particular subscore. While the 1984 survey indicated that the Listening Comprehension score was slightly favored by those responding, and the 1986 survey respondents indicated that Section 3, Vocabulary and Reading Comprehension, was the subscore most often emphasized, the 1989 survey found that the Listening Comprehension section score was used slightly more often than the other two section scores.

— Aside from test scores and subscores, what other information provided by the TOEFL test do admissions counselors consider when reviewing an application?

The TOEFL score report also contains personal information about the applicant that may help admissions officers. The 1989 survey asked institutions which of these criteria they used in their decision-making process. The 761 responding

institutions ranked the use of these factors as follows:

Native country	Used by 69%
Native language	Used by 65%
Date of birth	Used by 64%
Number of times TOEFL taken before	Used by 51%
Sex	Used by 45%
Degree planned	Used by 32%
Reason for taking TOEFL	Used by 25%

— Are there special TOEFL requirements for teaching assistantships?

As to a specific minimum TOEFL score requirement for teaching assistantship applicants, 121 of the 275 institutions responding (44 percent) indicated that they did require a minimum total TOEFL score. Of these institutions, 17 percent required a total TOEFL score range of 500-547, 68 percent required a range of 500-597, and 15 percent required a total TOEFL score of 600 or more. The TOEFL program recommends against the use of rigid cut-off scores.

— How are Test of Written English scores used?

At the time of the survey of TOEFL score users, 51 respondents currently recommended that their nonnative English-speaking applicants submit Test of Written English scores with their TOEFL scores. Only 13 currently required their nonnative English-speaking applicants to submit Test of Written English scores with their TOEFL scores. However, 101 institutions reported that they were considering recommending that their TOEFL applicants submit Test of Written English scores with their TOEFL scores; and 41 institutions were considering requiring TWE scores. Only 6 schools reported that they had completed in-house validity studies of their TWE score policies; 6 institutions reported that they were planning or conducting TWE validity studies.

Of the 552 respondents 311 (56 percent), reported that they administered writing tests to nonnative speakers of English upon enrollment. Of those who used the TWE score, most used it to help make admissions decisions in borderline cases (90 schools). Fifty schools reported that they used the TWE test score to help judge the validity of other tests, and 36 schools used TWE as an essential admissions criterion. Thirty-eight schools relied on the TWE score to help determine placement in English courses, and 28 schools used TWE scores to help determine placement in ESL courses.

— How are Test of Spoken English scores used?

Eighty-nine graduate institutions reported that the Test of Spoken English was required of teaching assistantship applicants. Of the 89, 67 reported that the TSE test was required of teaching assistantship applicants to all departments. Only 16 institutions reported that TSE was required for admission. Most schools relied on the overall comprehensibility score rather than the three TSE diagnostic scores reported for each TSE examinee.

Results of Survey of Institutional Policies

Institutional Policy	*Score Range*	*Number Responding*	*Percent Responding*	*Notes*
Begin academic work with no restrictions at undergraduate level	500-547 550-600	124 192	38% 59%	Of 324 respondents, only 6 allow students to begin work with no resitrctions when the TOEFL score is below 500. 3 of these were community colleges. 42 percent of undergraduate schools indicated a minimum of 525-550 as suitable for this category.
Begin academic work with no restrictions at graduate level	500-547 550-600	59 252	18% 77%	Only 1 indicated a requirement below 500. 49 indicated a minimum score of 600 or more. 63 percent of graduate schools indicated a minimum of 525-550 as suitable for this category.
Begin undergraduate work with no restrictions in selected fields not requiring high verbal ability (engineering, agriculture, etc.)	500-547 550-600	83 71	51% 43%	Of all 164 respondents, 35 percent reported a score of 500. Only 10 schools reported scores below 500.
Begin graduate work with no restrictions in selected fields, (engineering, math, chemistry, etc.)	500-547 550-600	47 92	33% 64%	57 percent indicated a range of 540-570 as the minimum for fields not requiring high verbal ability.
Begin academic work with initial limitation on academic load and with some supplemental English instruction	450-497 500-547 550-600	37 179 70	13% 62% 24%	Only 3 respondents indicated a score below 450.
Begin academic work if strong on all aspects except English proficiency; at least half-time ESL and corresponding reduction in academic load	450-497 500-547 550-600	53 102 34	28% 53% 18%	51 percent of 191 respondents indicated a minimum score between 480 and 520 as suitable for this requirement.
Be referred to full-time ESL program	450-497 500-547 550-600	111 107 12	42% 41% 5%	34 schools reported a score below 450; 10 of these were graduate schools, and 4 were community colleges.

Performance of Reference Groups on the TOEFL Test

Tables 2-6 present midpercentile rank data based on TOEFL scores obtained from July 1987 through June 1989. Table 2 gives the percentile ranks for section and total scaled scores for the total group. In Tables 3-6, examinees have been classified by their reason for taking the test (information supplied by them when they took it). These tables simply summarize the performance of self-selected groups of examinees who took TOEFL during the period specified.

Tables 7-8 present means and standard deviations of section and total scores, separately for male and female examinees tested between July 1987 and June 1989.

Table 9 gives the means of section and total scaled scores for all examinees, classified by native language. Table 10 gives the means of section and total scaled scores for nonnative English-speaking individuals, classified by native country. These tables may be useful in comparing the performance on the test of a particular student with that of other students from the same country and with that of students who speak the same language. It is important to point out that the data do not permit the generalization that there are fundamental differences in the ability of the various national and language groups to learn English or in the level of English proficiency they can attain. The tables are based simply on the performance of those examinees native to particular countries and languages who happened to take the TOEFL test.*

Because different selective factors may operate in different parts of the world to determine who takes TOEFL, the samples on which the tables are based are not necessarily representative of the student populations from which they came. In some countries, for example, virtually any high school, university, or graduate student who aspires to study in the United States or Canada may take the test. In other countries, government regulations permit only graduate students in particular areas of specialization, depending on national interests, to do so.

* More detailed information about the performance of selected major country and language groups is provided in TOEFL Research Report 11.

Table 2. Percentile Ranks for TOEFL Scores — Total Group*

Scaled Score	*Section 1 Listening Comprehension*	*Section 2 Structure and Written Expression*	*Section 3 Vocabulary and Reading Comprehension*	*Total Scaled Score*	*Percentile Rank*
68		98			
66	98	97	99	660	99
64	95	94	96	640	97
62	91	91	92	620	93
60	86	85	87	600	88
58	79	77	80	580	81
56	72	70	71	560	73
54	62	59	60	540	62
52	52	49	50	520	51
50	42	39	39	500	39
48	31	30	30	480	29
46	22	23	22	460	20
44	14	16	16	440	13
42	8	11	11	420	8
40	4	7	7	400	5
38	2	4	4	380	2
36	1	2	3	360	1
34	1	1	2	340	
32		1	1	320	
30				300	
Mean	51.8	51.9	51.6	Mean	518
S.D.	7.1	7.8	7.5	S.D.	67

* Based on the total group of 875,897 examinees tested from July 1987 through June 1989. (Total group includes 70,844 examinees not included in Tables 3-6 who did not indicate a "reason for taking TOEFL" or who indicated a reason other than those given in Tables 3-6.)

Table 3. Percentile Ranks for TOEFL Scores — Graduate-Level Students*

Scaled Score	*Section 1 Listening Comprehension*	*Section 2 Structure and Written Expression*	*Section 3 Vocabulary and Reading Comprehension*	*Total Scaled Score*	*Percentile Rank*
68		98			
66	98	96	98	660	99
64	96	93	95	640	96
62	92	88	90	620	92
60	87	82	83	600	87
58	81	72	74	580	79
56	73	63	64	560	69
54	64	51	52	540	57
52	53	40	40	520	44
50	42	31	30	500	32
48	31	23	21	480	22
46	21	16	15	460	14
44	13	11	10	440	9
42	7	7	7	420	5
40	4	4	4	400	3
38	2	2	2	380	1
36	1	1	1	360	1
34		1	1	340	
32				320	
30				300	
Mean	51.7	53.4	53.2	Mean	528
S.D.	6.8	7.4	7.0	S.D.	63

* Based on 423,366 examinees who, on their TOEFL answer sheets, indicated that they were applying for admission to colleges or universities as graduate students.

Table 4. Percentile Ranks for TOEFL Scores — Undergraduate-Level Students*

Scaled Score	Section 1 Listening Comprehension	Section 2 Structure and Written Expression	Section 3 Vocabulary and Reading Comprehension	Total Scaled Score	Percentile Rank
68	99	99			
66	98	97	99	660	99
64	94	95	97	640	97
62	90	92	94	620	93
60	83	88	91	600	89
58	76	81	85	580	83
56	68	76	78	560	76
54	59	67	70	540	67
52	49	57	60	520	57
50	39	48	49	500	45
48	29	38	39	480	35
46	21	29	30	460	25
44	13	21	22	440	16
42	8	14	15	420	10
40	4	9	10	400	6
38	2	5	6	380	3
36	1	3	4	360	1
34	1	2	2	340	
32		1	1	320	
30		1	1	300	
Mean	52.3	50.6	49.9	Mean	509
S.D.	7.3	7.9	7.6	S.D.	69

* Based on 342,929 examinees who, on their TOEFL answer sheets, indicated that they were applying for admission to colleges or universities as undergraduate students.

Table 5. Percentile Ranks for TOEFL Scores — Other Students*

Scaled Score	Section 1 Listening Comprehension	Section 2 Structure and Written Expression	Section 3 Vocabulary and Reading Comprehension	Total Scaled Score	Percentile Rank
66	99	99	99	660	
64	97	99	98	640	98
62	95	97	96	620	96
60	91	92	93	600	93
58	87	88	89	580	90
56	82	84	84	560	85
54	75	77	77	540	78
52	67	70	70	520	71
50	58	62	61	500	62
48	48	53	52	480	52
46	37	44	42	460	41
44	26	35	34	440	31
42	17	26	25	420	22
40	10	18	18	400	14
38	5	12	12	380	8
36	3	8	8	360	4
34	2	4	5	340	1
32		2	3	320	
30		2	1	300	
Mean	49.0	47.6	47.6	Mean	481
S.D.	7.4	8.3	8.0	S.D.	73

* Based on 18,219 examinees who, on their TOEFL answer sheets, indicated that they were applying for admission to schools other than colleges or universities, e.g., secondary schools.

Table 6. Percentile Ranks for TOEFL Scores — Applicants for Professional License*

Scaled Score	Section 1 Listening Comprehension	Section 2 Structure and Written Expression	Section 3 Vocabulary and Reading Comprehension	Total Scaled Score	Percentile Rank
68		98			
66	98	97	98	660	99
64	95	95	95	640	96
62	90	92	91	620	93
60	83	87	86	600	88
58	75	79	78	580	81
56	65	71	68	560	71
54	53	60	55	540	57
52	42	48	42	520	43
50	31	37	31	500	30
48	22	27	22	480	22
46	15	20	16	460	15
44	10	14	11	440	10
42	6	10	8	420	7
40	4	6	6	400	4
38	2	4	4	380	2
36	1	3	3	360	1
34	1	1	2	340	1
32		1	1	320	
30		1	1	300	
Mean	53.1	52.0	52.6	Mean	526
S.D.	6.9	7.5	7.1	S.D.	65

* Based on 20,499 examinees who, on their TOEFL answer sheets, indicated that they were taking TOEFL to become licensed to practice their professions in the United States or Canada.

Table 7. Means and Standard Deviations for TOEFL Section and Total Scores, Males*

Group	Number	Section 1 Listening Comprehension	Section 2 Structure and Written Expression	Section 3 Vocabulary and Reading Comprehension	Total Scaled Score
Total Group†	531,829	51.5 7.1	52.0 7.9	51.9 7.6	518 68
Graduate Level	280,686	51.5 6.8	53.6 7.3	53.6 6.9	529 63
Undergraduate Level	191,980	52.1 7.4	50.3 8.0	49.9 7.7	508 71
Other Students	9,134	49.0 7.6	47.9 8.7	48.2 8.4	484 76
Applicants for Professional License	9,791	53.1 7.2	52.6 7.7	53.5 7.4	531 68

* Based on examinees tested from July 1987 through June 1989 who responded to sex group membership on answer sheets.

† Total group includes 40,238 examinees not included in the four succeeding groups.

Table 8. Means and Standard Deviations for TOEFL Section and Total Scores, Females*

Group	Number	Section 1 Listening Comprehension	Section 2 Structure and Written Expression	Section 3 Vocabulary and Reading Comprehension	Total Scaled Score
Total Group†	342,047	52.2 7.1	51.7 7.6	51.1 7.3	517 67
Graduate Level	141,583	52.1 6.8	52.9 7.4	52.5 6.9	525 64
Undergraduate Level	150,297	52.6 7.2	50.9 7.7	50.0 7.3	512 68
Other Students	9,051	49.0 7.2	47.4 7.9	47.1 7.6	478 70
Applicants for Professional License	10,672	53.2 6.6	51.4 7.2	51.8 6.8	521 63

* Based on examinees tested from July 1987 through June 1989 who responded to sex group membership on answer sheets.

† Total group includes 30,444 examinees not included in the four succeeding groups.

Statistical Characteristics of the Test

Level of Difficulty

It is generally agreed by measurement specialists that the TOEFL test will provide the best measurement in the critical score range of about 450 to 550 when the test is of middle difficulty. That is, the average raw score for each section as a percentage of the maximum possible raw score should be about midway between a perfect score and the score that would be expected if the answer to each question were marked at random. For TOEFL, that value is 62.5 percent of the maximum possible score for a section. The mean percent correct for the sections for the twenty-four different forms administered between July 1987 and June 1989 all fall within 54.4 percent and 78.9 percent of the maximum possible score. For Listening Comprehension, the average percent correct ranges from 54.4 to 72.4 percent, with a mean of 64.4 percent. For Structure and Written Expression, the values range from 66.8 to 78.9 percent, with a mean percent correct of 71.7. For Vocabulary and Reading Comprehension, the values range from 60.5 to 75.0 percent, with a mean percent correct of 65.9.

Percent correct, as a measure of difficulty, depends both on the inherent difficulty of the test and on the ability level of the group of examinees that took the test. Both factors are of concern in determining whether the test is properly matched to the ability level of the examinees. However, for the scaled scores that are reported to examinees and institutions, the effect of the differences in difficulty level among the various forms of the test is removed, or adjusted for, by a statistical process called score equating. (See "TOEFL Scores," page 11.)

To equate scores, the TOEFL program employs an equating method known as item response theory. This method utilizes item-response data obtained in previous administrations to place items on a common scale. To accomplish this, the statistical parameters of items scaled (calibrated) at the pretest phase are utilized when the items are administered in a final form. For test forms that include items that have not been previously scaled, an external anchor of previously calibrated items is used as a basis for equating the unscaled items to a uniform scale. For additional information about item response theory, see Lord (1980) or McKinley (1989).

Comparability of scores across TWE topics is established through carefully controlled procedures for topic development, pretesting, pretest analysis, and postadministration reading. Because guessing is not a factor for TWE, only the mean scores are reported to summarize performance on the essay question. (TWE scores range from 1-6.) For the March 1989 administration, the mean score on the TWE for the total group of examinees was 3.66, with a standard deviation of 0.95. For the May 1989 administration, the mean score was 3.78 with a standard deviation of 0.93.

Adequacy of Time Allowed

While no single statistic has been widely accepted as a measure of the adequacy of time allowed for a separately timed section, two rules of thumb are used at ETS: (1) 80 percent of the group ought to be able to finish about every question in each section, and (2) 75 percent of the questions in a section ought to be completed by almost all of the group. The Listening Comprehension section of TOEFL is paced by

Table 9. TOEFL Total and Section Score Means — All Examinees Classified by Native Language*

(Based on 875,897 students seeking admission to institutions in the United States or Canada who took TOEFL from July 1987 through June 1989)**

Native Language	Number of Examinees	Listening Comprehension	Structure and Written Expression	Vocabulary and Reading Comprehension	Total Score Mean	Total Number of Examinees January 1979-June 1989†
Afrikaans	161	56	56	54	554	671
Akan	926	53	58	56	557	4,345
Amharic	2,323	51	50	49	500	8,244
Arabic	48,880	51	48	46	480	284,104
Armenian	587	55	53	51	530	4,278
Assamese	193	55	58	58	570	792
Aymara	0	*	*	*	*	32
Baluchi	1	*	*	*	*	161
Bambara	0	*	*	*	*	527
Basque (Euskara)	101	55	55	56	553	241
Bemba	103	56	60	57	579	365
Bengali	10,059	50	52	52	512	33,826
Berber	88	49	50	49	492	769
Bihari	1	*	*	*	*	138
Bulgarian	144	55	55	54	548	485
Burmese	840	51	51	51	513	2,726
Catalan (Provencal)	442	55	56	57	561	1,224
Cebuano	674	58	57	57	571	3,633
Chichewa	188	55	59	56	567	495
Chinese	254,557	51	52	52	509	931,046
Czech	334	57	55	55	558	1,118
Danish	1,137	62	59	57	593	3,853
Dutch	3,670	62	59	58	598	14,488
Efik	162	52	54	53	529	2,924
English	7,598	59	58	58	584	28,536
Estonian	0	*	*	*	*	31
Ewe	365	52	58	55	551	1,493
Farsi (Persian)	8,096	54	50	49	509	71,274
Fijian	52	56	55	52	542	406
Finnish	940	60	58	57	584	3,317
French	23,907	54	55	56	551	70,599
Fulani (Fula)	101	48	51	50	497	567
Ga	268	55	59	57	569	1,624
Galla (Gallinya)	67	49	52	51	510	276
Ganda	2	*	*	*	*	2
Georgian	0	*	*	*	*	56
German	15,769	60	59	57	586	44,110
Gilbertese	0	*	*	*	*	3
Greek	16,284	54	53	50	526	75,465
Guarani	20	53	48	50	500	116
Gujarati	6,535	54	54	54	540	27,257
Hausa	230	51	54	52	526	2,712
Hebrew	4,375	59	54	53	551	17,393
Hindi	10,488	57	60	59	587	34,386
Hungarian	657	57	55	55	558	1,803
Ibo (Igbo)	1,450	52	55	53	532	30,393
Icelandic	939	60	55	54	563	3,178
Ilocano (Ilokano)	594	54	53	53	530	3,440
Indonesian	23,559	51	49	49	496	80,648
Italian	4,259	56	57	58	569	14,673
Japanese	152,884	49	50	49	485	346,084
Javanese	1,153	50	49	48	491	4,297
Kannada (Kanarese)	2,669	55	59	59	577	8,558
Kashmiri	135	56	60	58	582	542
Kazakh	0	*	*	*	*	4
Khalkha Mongolian	133	50	51	51	505	339
Khmer	472	52	47	47	487	1,515
Kikongo	0	*	*	*	*	305
Kikuyu	628	55	59	56	563	2,272
Korean	67,371	48	52	53	505	242,459
Kurdish	68	53	49	48	499	394
Kusaiean	114	47	46	44	454	440
Lao	269	53	47	48	491	1,205
Latvian	0	*	*	*	*	31
Lingala	99	48	49	48	483	342
Lithuanian	0	*	*	*	*	60
Luba-Lulua	89	48	52	50	502	327
Luganda	0	*	*	*	*	791
Luo	334	53	58	55	555	1,053
Macedonian	1	*	*	*	*	80
Malagasy	69	50	53	53	523	176
Malay (Bahasa)	15,557	56	53	53	540	105,738
Malayalam	2,423	56	59	59	579	9,092
Malinke (Mandingo)	167	49	52	50	502	670
Maltese	40	60	62	61	607	161
Marathi	2,456	57	60	60	590	8,386
Marshallese	218	48	45	42	451	950
Mende	102	51	56	54	537	541
Nepali	1,143	52	54	54	535	3,393
Norwegian	3,609	60	57	55	573	11,921
Onya	447	52	57	58	556	1,586
Palauan	305	52	47	46	482	1,471
Pidgin	17	56	56	54	552	35
Polish	2,476	56	54	54	547	8,298
Ponapean	90	53	52	49	511	878
Portuguese	8,828	53	53	54	534	28,637
Punjabi	3,876	53	55	54	540	12,713
Pushtu	763	50	51	51	504	2,587
Quechua	0	*	*	*	*	19
Rajasthani	3	*	*	*	*	306
Romanian	586	57	56	56	564	2,658
Russian	953	55	54	53	540	9,216
Samoan	432	55	51	49	516	4,799
Sangho	0	*	*	*	*	40
Sepedi	0	*	*	*	*	121
Serbo-Croatian	1,432	57	54	54	551	3,877
Sesotho	127	55	59	54	563	364
Setswana	305	55	58	54	557	1,002
Shona	291	57	60	57	583	969
Sindhi	883	55	56	55	554	2,923
Sinhalese	2,728	55	55	55	553	8,142
Siswati	147	56	59	55	568	482
Slovak	142	57	55	55	558	508
Slovene	1	*	*	*	*	191
Somali	565	50	48	48	484	2,824
Spanish	44,368	54	52	54	534	220,273
Sundanese	136	49	49	49	490	409
Swahili	1,194	53	56	54	542	4,228
Swedish	2,203	62	58	57	591	8,259
Tagalog	7,581	58	57	57	573	35,846
Tahitian	0	*	*	*	*	46
Tamil	7,849	56	59	59	578	26,611
Tatar	0	*	*	*	*	25
Telugu	6,933	53	56	57	553	20,802
Thai	24,939	49	49	49	489	105,621
Tibetan	76	55	55	55	547	211
Tigrinya	445	52	52	50	512	1,677
Tongan	53	50	46	46	475	528
Trukese	122	46	45	42	443	733
Tulu	144	57	58	58	577	416
Turkish	9,841	52	52	50	516	28,472
Twi-Fante	513	55	60	58	575	2,599
Uighur	0	*	*	*	*	24
Ukrainian	0	*	*	*	*	194
Ulithian	616	53	55	55	544	962
Urdu	15,579	52	52	51	518	52,368
Vietnamese	4,380	53	51	51	513	28,926
Visayan	2	*	*	*	*	1,656
Wolof	338	50	52	50	508	1,180
Yapese	147	56	52	52	531	496
Yi	0	*	*	*	*	1
Yoruba	1,202	54	56	55	548	18,376
Zhuang	9	*	*	*	*	45
Zulu	157	55	58	55	560	725

* Because of the unreliability of statistics based on small samples, means are not reported for subgroups of less than 15 for a total of 20 examinees.

** Includes 17,345 students who did not report their native language.

† Does not include students who took TOEFL from July 1986 through June 1987.

a recording; thus, every question is presented to every examinee and the criteria for speededness do not apply. There is no way of determining from the data available the adequacy of the time allowed for responding to each question.

For Sections 2 and 3 of the twenty-four forms administered between July 1987 and June 1989, at least 80 percent of each group of examinees were able to complete **all** the questions in each section, and the three-quarter point in the sections was reached by 99.5 to 99.9 percent. Thus, one may reasonably conclude that speed is not an important factor in TOEFL scores by these criteria.

These speededness criteria are not applicable to the TWE.

Table 10. TOEFL Total and Section Score Means — Nonnative English-Speaking Examinees Classified by Geographic Region and Native Country*

(Based on 875,897 students seeking admission to institutions in the United States or Canada who took TOEFL from July 1987 through June 1989)**

Geographic Region and Native Country	Number of Examinees	Listening Comprehension	Structure and Written Expression	Vocabulary and Reading Comprehension	Total Score Mean	Total Number of Examinees January 1979-June 1989†
AFRICA						
Algeria	670	50	50	50	500	4,666
Angola	34	55	54	54	543	112
Benin	86	48	52	51	501	223
Botswana	252	55	58	54	555	837
Burkina Faso (Upper Volta)	58	50	54	52	518	201
Burundi	90	49	53	51	511	273
Cameroon	634	50	53	50	509	2,816
Cape Verde	55	52	50	49	506	244
Central African Republic	57	45	48	47	465	136
Chad	23	48	50	47	482	60
Comoros	33	48	50	47	481	37
Congo	45	46	49	48	478	149
Djibouti	51	48	47	46	467	192
Egypt	6,775	52	51	49	505	32,305
Equatorial Guinea	8	*	*	*	*	25
Ethiopia	2,868	51	51	49	503	10,157
Gabon	55	50	51	50	503	183
Gambia	228	51	54	51	520	758
Ghana	2,019	54	59	57	569	9,913
Guinea	154	49	51	49	496	512
Guinea, Bissau	30	47	45	47	465	73
Ivory Coast	445	49	50	49	495	1,990
Kenya	2,150	54	58	55	559	8,161
Lesotho	80	55	60	54	562	299
Liberia	178	51	54	51	517	1,251
Libya	371	53	49	47	496	7,450
Madagascar	123	52	54	54	537	341
Malawi	166	55	59	56	567	468
Mali	125	50	53	52	515	678
Mauritania	80	51	52	51	516	186
Morocco	2,210	50	51	49	500	6,891
Mozambique	75	53	53	54	533	159
Namibia	40	47	50	49	488	40
Niger	48	50	54	51	517	243
Nigeria	3,330	53	55	54	540	62,103
Rwanda	69	48	53	51	505	201
Sao Tome and Principe	25	50	49	48	490	45
Senegal	360	49	51	50	500	1,107
Seychelles	13	*	*	*	*	56
Sierra Leone	204	52	57	54	546	1,277
Somalia	517	49	48	47	482	2,704
South Africa	452	56	58	56	567	2,088
Sudan	1,150	48	50	48	487	5,040
Swaziland	154	56	59	55	566	565
Tanzania	627	53	56	54	545	2,951
Togo	136	47	51	50	495	417
Tunisia	1,028	50	52	51	510	4,197
Uganda	438	55	60	56	570	1,591
Zaire	496	49	51	50	499	1,668
Zambia	235	57	60	57	583	1,314
Zimbabwe	346	57	60	57	581	1,248
AMERICAS						
Antigua and Barbuda	0	*	*	*	*	1
Argentina	2,552	56	56	57	560	9,853
Bahamas	9	*	*	*	*	38
Barbados	0	*	*	*	*	12
Belize	11	55	52	51	528	40
Bermuda	0	*	*	*	*	8
Bolivia	1,068	54	51	52	523	5,313
Brazil	8,032	53	52	54	532	25,778
Canada	1,754	57	53	54	547	6,200
Cayman Islands	0	*	*	*	*	1
Chile	1,574	53	51	55	532	7,681
Colombia	4,584	53	51	54	527	24,263
Costa Rica	1,149	55	53	55	541	4,226
Cuba	308	56	54	57	557	3031
Dominica	7	*	*	*	*	8
Dominican Republic	1,025	54	51	53	526	4,138
Ecuador	1,260	54	51	53	523	6,072
El Salvador	1,245	54	51	52	523	6,244
French Guiana	1	*	*	*	*	1
Grenada	2	*	*	*	*	6
Guadaloupe	3	*	*	*	*	3
Guatemala	998	56	53	55	546	4,541
Guyana	3	*	*	*	*	137
Haiti	1,635	51	51	50	505	7,419
Honduras	1,224	54	52	53	527	4,535
Jamaica	10	*	*	*	*	150
Martinique	1	*	*	*	*	1
Mexico	7,146	55	53	55	539	34,507
Netherlands Antilles	601	59	55	54	558	2,648
Nicaragua	517	55	52	55	539	3,233
Panama	2,334	53	50	51	516	8,393
Paraguay	241	55	53	54	540	748
Peru	3,466	54	52	54	534	13,529
Puerto Rico	4,429	54	52	54	530	14,156
St. Christopher and Nevis	1	*	*	*	*	1
St. Lucia	4	*	*	*	*	4
Suriname	157	58	57	54	562	965
Trinidad and Tobago	7	*	*	*	*	136
United States	1,633	59	56	55	566	6,010
Uruguay	342	55	55	57	558	1,335
Venezuela	2,299	54	51	53	525	49,679
Virgin Islands	2	*	*	*	*	12
ASIA						
Afghanistan	398	53	50	49	505	1,641
Bangladesh	7,052	48	48	49	482	22,800
Bhutan	62	55	56	56	557	212
Brunei	162	55	52	52	530	1,210
Cambodia (Kampuchea)	567	52	47	48	490	1,649
China, People's Republic	78,894	51	54	53	521	131,409
Hong Kong	71,596	52	50	51	506	334,982
India	44,487	56	58	58	571	151,724
Indonesia	23,971	51	49	49	496	83,916
Japan	154,609	49	50	49	485	350,110
Kiribati	6	*	*	*	*	29
Korea	67,834	48	52	53	505	244,823
Laos	299	53	48	49	499	1,175
Macau (Macao 88.89)	1,796	52	51	50	506	7,016
Malaysia	23,414	55	52	53	535	156,902
Mauritius	278	58	59	59	586	1,118
Mongolia	18	55	52	53	534	101
Myanmar, Union of	875	51	51	51	513	2,894
Nepal	1,136	52	54	54	532	3,243
Pakistan	17,475	52	51	51	513	57,246
Philippines	9,524	58	57	57	569	46,384
Singapore	4,728	60	57	58	583	27,369
Yemen	1,388	49	47	44	465	5,408
Sri Lanka	4,419	54	54	54	541	14,063
Taiwan	84,098	50	51	51	505	375,426
Thailand	25,027	49	49	49	489	105,895
Vietnam	4,810	53	51	51	514	31,209
EUROPE						
Albania	8	*	*	*	*	34
Andorra	25	51	51	51	508	52
Austria	1,051	60	60	58	594	2,706
Azores	10	*	*	*	*	10
Belgium	1,511	60	59	59	592	6,175
Bulgaria	118	56	55	55	553	347
Cyprus	4,459	53	51	48	509	23,588
Czechoslovakia	464	57	55	55	558	1,528
Denmark	1,174	62	59	57	593	3,907
Finland	1,015	61	59	57	586	3,520
France	17,694	54	55	56	554	44,779
East Germany	314	58	56	54	559	919
West Germany	11,912	60	59	58	588	33,301
Great Britain	201	57	56	55	563	990
Greece	12,282	55	53	51	532	53,175
Greenland	5	*	*	*	*	29
Hungary	587	57	55	55	559	1,529
Iceland	953	60	55	54	563	3,209
Ireland	20	57	49	52	527	90
Italy	4,122	56	57	58	569	14,210
Liechtenstein	32	59	56	53	561	63
Luxembourg	89	60	61	60	604	345
Madeira Islands	4	*	*	*	*	13
Maldives	10	*	*	*	*	40
Malta-Malta	74	55	56	56	556	224
Monaco	47	55	55	54	547	124
Netherlands	2,377	62	60	59	605	8,953
Norway	3,658	60	57	55	572	13,017
Poland	2,483	56	54	54	547	8,439
Portugal	612	58	56	56	566	2,546
Romania	655	57	56	56	564	2,932
San Marino	2	*	*	*	*	4
Soviet Union	1,008	55	53	53	538	9,460
Spain	7,001	55	55	56	551	15,545
Sweden	2,174	62	58	57	589	8,140
Switzerland	3,411	58	57	56	569	9,627
Turkey	9,511	52	52	50	515	27,469
Vatican City	10	*	*	*	*	34
Yugoslavia	1,609	57	54	54	551	4,284
MIDDLE EAST						
Bahrain	585	52	47	45	490	3,698
Iran	8,252	54	50	48	508	74,178
Iraq	777	53	49	48	498	9,355
Israel	4,886	58	53	52	544	20,065
Jordan	9,425	50	47	45	472	59,091
Kuwait	6,088	50	44	42	451	22,598
Lebanon	5,864	54	52	50	521	45,026
Oman	741	51	45	44	467	3,590
Qatar	598	50	44	42	451	3,080
Saudi Arabia	6,891	50	46	44	467	49,201
Syria	2,961	53	49	48	500	16,817
United Arab Emirates	1,711	50	45	43	462	7,541
PACIFIC REGION						
American Samoa	470	54	50	48	508	5,783
Australia	75	54	50	50	511	502
Caroline Islands	604	50	47	45	477	3,577
Fiji	119	57	57	54	560	840
Guam	12	*	*	*	*	25
Mariana Islands	29	57	52	51	531	249
Marshall Islands	216	48	45	42	452	937
Nauru	3	*	*	*	*	26
New Caledonia	1	*	*	*	*	1
New Zealand	8	*	*	*	*	33
Papua New Guinea	15	56	54	51	538	77
Solomon Islands	16	54	52	50	521	35
Tahiti	101	53	52	52	526	115
Tonga	68	50	47	47	479	534
Tuvalu	3	*	*	*	*	24
Vanuatu	5	*	*	*	*	5
Western Samoa	20	57	54	51	538	105

* Because of the unreliability of statistics based on small samples, means are not reported for subgroups of less than 15 for a total of 169 examinees.

** Includes 22,208 students who did not report their country of birth or who reported English as their native language.

† Total does not include students who took TOEFL from July 1986 through June 1987.

Reliabilities and the Standard Error of Measurement

The reliability of a test is the extent to which it yields consistent results. The reliability estimates for sections are measures of internal consistency or item homogeneity within each section. This estimate does not take into account other forms of variation such as form-to-form variation or day-to-day variation which, for practical reasons, cannot be estimated. Table 11 gives average internal consistency reliabilities of the scaled scores for the three multiple-choice sections and the total test for twenty-four forms administered between July 1987 and June 1989. For Section 1, the average of the reliabilities for the twenty-four forms was .86. For Section 2, the average of the reliabilities was .85. For Section 3, the average of the reliabilities was .89. For the total test, the average of the reliabilities was .95.

Table 11. Reliabilities and Standard Errors of Measurement*

Section	Reliability	SE of Measurement
1. Listening Comprehension	.86	2.5
2. Structure and Written Expression	.85	2.8
3. Vocabulary and Reading Comprehension	.89	2.4
Total Score	.95	14.8

* Averaged over twenty four forms administered between July 1987 and June 1989. Based only on examinees tested in the United States and Canada

The standard error of measurement is an estimate of the probable extent of the error inherent in a test score due to the imperfect precision of the measurement process. If it were possible to administer the same or closely parallel forms of the test to a given individual many times, the resulting scores would be expected to vary somewhat due to errors of measurement. The score the individual would achieve on the test if there were no errors of measurement is called the "true score." The observed scores are normally distributed around this true score. The standard deviation of that distribution is the standard error of measurement.

The standard error of measurement is expressed in the same units as the reported score, which makes it very useful in interpreting the scores of individuals. Table 11 shows that the standard error of measurement for Section 1 is 2.5 points; for Section 2, 2.8 points; for Section 3, 2.4 points; and for the total score, 14.8 points. That is, for Section 1, approximately two-thirds of the examinees have true scores within 2.5 points of their reported scores, and about 95 percent have true scores within 5.0 points of their reported scores. For Section 2, these intervals are slightly wider. Approximately two-thirds of the examinees have true scores extending 2.8 points on either side of their reported scores, and about 95 percent have true scores within 5.6 points of their reported scores. For Section 3, approximately two-thirds of the examinees have true scores within 2.4 points of their reported scores, while 95 percent have true scores within 4.8 points of their reported scores. For the total score, the corresponding intervals extend 14.8 points on either side of the reported score for two-thirds of the examinees and 29.6 points on either side of the reported score for 95 percent of the examinees.

In comparing total scores for two examinees, one should not conclude that one score represents a significantly higher level of proficiency in English than another score unless there is a difference of at least 30 points between them. In comparing section scores for two persons, the difference should be at least 5 points for Section 1, at least 6 points for Section 2, and at least 5 points for Section 3.

The reliability of the TWE is evaluated by examining the agreement between the two ratings assigned to each essay paper. As a measure of scorer agreement, the product-moment correlation between first and second ratings is computed. For both the March 1989 and May 1989 administrations, the correlation was .76. In addition to this correlation, coefficient alpha, which is a measure of agreement between ratings taking into account that there are two ratings per essay, is computed. Coefficient alpha for both the March 1989 and May 1989 administrations was .87

Reliability of Gain Scores

The most obvious way to measure a change in score from one administration to another is to note the simple difference between the two scores. Unfortunately, this procedure has some limitations (see, for example, Linn and Slinde, 1977, pages 121-150). The reliability of the difference of two scores is typically less than the reliability of either score alone. Other things being equal, the reliability of the difference decreases as the correlation between pretest and posttest increases. For example, if the reliability of both pretest and posttest is about .90 and if the standard deviations of the scores are assumed to be equal, the reliability of the gain scores decreases from .80 to .50 as the correlation between pre- and posttest increases from .50 to .80. If the correlation is as high as the reliability, the reliability of the gain scores is zero.

The analysis of data obtained from a group of students at San Francisco State University indicates that score gains decrease as a function of proficiency level at the time of initial testing (Swinton, 1983). For this group, student scores were obtained at the start of an intensive English language program and at its completion thirteen weeks later. Students whose initial scores were in the 353-400 range showed an average gain of 61 points while students whose initial scores were in the 453-500 range showed an average gain of 42 points.

The attribution of gains requires caution, since they may reflect increased language proficiency, a practice effect, and regression toward the mean. In the case of this group, an attempt was made to remove the effects of regression and practice by administering another form of TOEFL one week after the pretest. Initial scores in the 353-400 range increased about 20 points on the retest, while initial scores in the 453-500 range improved about 17 points on the retest. The greater part of these gains can be attributed to the effects of practice and regression, although a small part may reflect the effect of one week of instruction.

Subtracting the retest gain (20 points) from the posttest gain (61 points), it was possible to determine that, within this sample, students with initial scores in the 353-400 range showed a real gain on the TOEFL of 41 points during thirteen weeks of instruction. Similarly, students in the 453-500 initial score range showed a 25 point gain in real language proficiency after adjusting for the effects of practice and regres-

sion. Thus, the lower the initial score, the greater will be the probable gain over a fixed period of instruction. Other factors, such as the nature of the instructional program, will affect gain scores also.

The TOEFL program has published a manual (Swinton, 1983) that describes a methodology suitable for conducting local gains studies. University-affiliated and private English language institutes may wish to conduct gains studies with their own students to determine the amount of time that is ordinarily required to improve from one score level to another.

Intercorrelations Among Scores

The three multiple-choice sections of the TOEFL test are designed to measure different skills within the general domain of English proficiency. It is commonly recognized that these skills are interrelated; persons who are highly proficient in one area tend to be proficient in the other areas as well. If this relationship were perfect, there would be no need to report scores for each section. The scores would represent the same information repeated several times, rather than different aspects of language proficiency.

Table 12 gives the correlation coefficients measuring the extent of the relationships among the three sections and with the total test score. A correlation coefficient of 1.0 would indicate a perfect relationship between the two scores, and 0.0 would indicate a total lack of relationship. The table shows average correlations over the twenty-four forms administered between July 1987 and June 1989. Correlations between the section scores and the total score are spuriously high because the section scores are included in the total. The observed correlations indicate that there is a fairly strong relationship among the skills tested by the three multiple-choice sections of the test, but that the section scores provide some unique information.

Table 12. Intercorrelations Among the Scores*

Section	1	2	3	Total
1. Listening Comprehension	—	.68	.70	.87
2. Structure and Written Expression	.68	—	.79	.92
3. Vocabulary and Reading Comprehension	.70	.79	—	.93
Total Score	.87	.92	.93	—

* Averaged over twenty-four forms administered between July 1987 and June 1989. Based only on examinees tested in the United States and Canada.

Correlations are also obtained between the TWE scores and the TOEFL section and total scores. For the March 1989 administration, the correlation between TWE scores and Section 1 scaled scores was .65. The correlation between TWE scores and Section 2 scaled scores was .64, and, for Section 3 scores and TWE scores, the correlation was .62. The TOEFL-TWE total score correlation was .67. For the May 1989 administration, the correlation between TWE scores and Section 1 scaled scores was .63. The correlation between TWE scores and Section 2 scaled scores was .64, and, for Section 3 scaled scores and TWE scores, the correlation was .62. The TOEFL-TWE total score correlation was .67. From these data, it can be concluded that a moderate relationship exists between skills tested by the three multiple-choice sections of TOEFL and the TWE and that the TWE measures abilities distinct from those measured by TOEFL.

Validity

In addition to being reliable, there should be an indication that a test is valid, that is, that it actually measures what it is intended to measure. For example, a test of basic mathematical skills that yielded very consistent scores would be considered reliable. But if those scores showed little relationship to students' performance in basic mathematics courses, the validity of the test would be questionable. This would be particularly true if the scores did show a strong relationship to the students' performance in other areas, such as language or social studies. For TOEFL, the question of validity relates to how well the test measures a person's proficiency in English as a second or foreign language.

Establishing the validity of a test is admittedly one of the most difficult tasks facing those who design the test. For this reason, validity is usually confirmed by analyzing the test from a number of perspectives.

Although researchers have stated definitions for many different types of validity, it is generally recognized that validity refers to the usefulness of inferences made from test scores (APA, 1985). To support inferences, validation should include several types of evidence, e.g., content-related, criterion-related, and construct-related. The nature of the evidence should depend on the specific inference or use of the test.

To establish content-related evidence, one must demonstrate that the content and behavior exhibited on a test constitute an adequate sample of the content and behaviors of the subject or field tested. Criterion-related evidence of validity applies when one wishes to draw a relationship between a score on some other variable, called a criterion, and a score on the test under consideration. Two basic types of criterion-related evidence refer to the traits or skills that are related to variation in test scores. For a test that reports three section scores and a writing score, such as TOEFL-TWE, research should provide evidence that the test measures four different constructs. Such research should also support the label assigned to each construct in the name of each section. Of the three kinds of validity evidence, content-related evidence is established by examining the content of the test, whereas criterion-related and construct-related evidence frequently involve judgments based on statistical relationships.

Content-related evidence for the TOEFL is the responsibility of the TOEFL Committee of Examiners (see page 6), which has developed a comprehensive list of specifications for items appearing in the different sections of the test. The specifications identify the aspects of English that are to be tested and describe appropriate techniques for testing them. The specifications are continually reviewed and revised as appropriate to ensure that the test reflects both current English usage and current theory as to the nature of second language proficiency.

The TOEFL Research Committee (see page 6) has undertaken a program to systematically analyze the content of the test through rigorous research methodologies. A study by Duran, Canale, Penfield, Stansfield, and Liskin-Gasparro (1985) analyzed one form of TOEFL from several different frameworks related to contemporary ideas about aspects of communicative competence. These frameworks take into account the grammatical, sociolinguistic, and discourse competencies required to answer TOEFL items correctly.

While the competencies and the degree to which TOEFL measures them vary considerably across sections, the results indicate that successful performance on the test requires a wide range of competencies. Although the authors' findings are too numerous to cite in detail, a few of them are described below.

In general, it was found that the range and complexity of skills required on various sections of the TOEFL test are directly related to the amount of language used and to the semantic and textual complexity of TOEFL items. The minitalks and extended conversations in Section 1 and the reading comprehension passages in Section 3 are the only parts of the test that involve multiple items based on a single context. Because the context of these items is more fully developed than the context of items in the rest of the test, these items are more natural and require a greater number of communicative skills of examinees.

Section 1 is the only section of the test requiring skills that involve the recognition of pronunciation. This section attempts to simulate actual listening tasks. While clear enunciation helps ensure that examinees have a maximum chance to recognize the meaning of utterances, it does not require examinees to utilize speech recognition skills that deviate far from the lexical and grammatical structure of formal academic language. A similar type of speech, alluded to as "foreigner talk," has been found to occur when native speakers make a special effort to communicate with nonnative speakers. Although Part A of Section 1 seems to focus principally on recognition of phonetic, grammatical, and lexical aspects of spoken English, the language of the stimulus material makes use of several speech act functions. Over half of the statement items in Part A give information simply and directly, while the remaining may indirectly advise, invite, seek approval, make a suggestion, and so forth. Over half of the statement items contain idiomatic expressions, which are usually noun or verb compounds.

According to Duran et al. (1985), idiomatic expressions are even more prevalent in the dialogues that make up Part B of Section 1. They occurred in approximately two-thirds of the stimuli for these items. In contrast to items in Part A, most dialogue items are not direct requests for, or offers of, information. Rather, such items reflect several other direct and indirect communicative purposes. Some of the indirectly expressed messages in dialogue stimuli are request for help, promise, complaint, slight insult, suggestion, impatience, annoyance, advice, and denial. About 20 percent of the statements made by the second speaker reflect a direct or an indirect response to the request of the first speaker. Another 20 percent reflect a modification of the first speaker's comment. Other dialogues reflect elaboration of information, clarification, and suggestions offered by the second speaker in reaction to the first speaker's statement or comment.

Part C of Section 1 consists of questions based on several minitalks and extended conversations. Duran et al. found that Part C displays a formal style of speech that closely approximates expository prose in its vocabulary, sentence formation, and discourse structure. Idiomatic forms are less prevalent in this part of Section 1. Nonetheless, due to the length and complexity of these texts, many more competencies are required to understand stimulus passages. Sentence formation aural recognition skills include simple sentence word order and compound and complex sentence word order with embedded subordination. The topic of stimulus passages is relevant to academic life or classroom content, and the passages are drawn verbatim from academic texts. In approximately half the items in this part, the relationship between test questions and correct response options is elliptical. Ellipsis, or the deletion of mutually understood information, is typical of natural communication.

Information regarding the perceptions of college faculty of the validity of the Listening Comprehension section is available in *A Survey of Academic Demands Related to Listening Skills* (Powers, 1985). Powers found that "the three kinds of listening comprehension questions used currently in the TOEFL were rated (by faculty) as being among the most appropriate of those considered" (p.14).

Section 2 of the TOEFL test, Structure and Written Expression, requires more diverse orthographic recognition skills than does Section 1. Duran et al. found that capitalization-marking formal identity in English, such as the proper names of people, places, books, historical events, dates, and nationalities, occurred in almost 50 percent of the items. The most essential competencies necessary for both parts of Section 2 are concerned with the recognition of vocabulary and word formation, along with phrase, clause, and sentence formation. Idiomatic expressions occur frequently, and function words and inflected forms of content words test a variety of grammar-based concepts. Some of these concepts include location, direction, quantity, time, possession, and comparison. Many of the items in this section refer to classroom content that is typical of secondary school or college courses in North America. Discourse competencies required in Section 2 include the recognition of lexical and sentence-level cohesive devices, such as pronouns, clausal reduction, parallel structure, and conjoiners binding semantic relationships between clauses.

According to Duran et al., Section 3, Vocabulary and Reading Comprehension, is more oriented toward the conveyance of meaning than is Section 2. The two parts in this section differ markedly in the length of item stimuli. This explains why the reading comprehension part tests a greater variety of competencies than does the vocabulary part. Because vocabulary items are shorter, they test competencies that are more similar to those tested in Section 2.

The stimulus portion of vocabulary items has been found to manifest idiomatic vocabulary frequently, especially noun compounds, and function words indicating special relationship and quantity. Although most stimulus sentences exhibit the subject-verb-object word order, passive constructions are frequently used as well. Sentence-level cohesion devices are also noticeable. Although the stimulus sentences used in vocabulary items are fairly short, this helps to ensure that examinees will not infer the meaning of the tested word solely on the basis of the words in the stimulus. Thus, the use of short stimuli maintains the validity of this part as a measure of vocabulary while integrating sentence level features into the item.

In their content analysis of one form of the test, Duran et al. judged that the passages in the reading comprehension part form a rich sample of language for examinees. The passages

invoke an almost exhaustive array of grammatical competencies and many discourse competencies. The passages exhibit complex sentence patterns known to be typical of advanced expository prose, such as that found in college textbooks and technical writing. The passages frequently rely on cohesive devices to bind interclausal semantic or propositional relationships, although they do not always exhibit the planned organization of ideas that one might expect to find in formal rhetoric. The questions for reading comprehension passages often focus on literal meaning, such as main ideas and details. In other questions, the examinee must infer details not explicitly stated in the passage, such as probable consequences, or infer the writer's viewpoint, attitude, or tone.

For additional information related to the content validity of the test, see "What TOEFL Is Like" on page 7.

In 1984, the TOEFL program held an invitational conference to discuss the content validity of the current test. The conference brought together some two dozen specialists in the testing of English as a second language. The papers presented at the conference are available in *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference* (Stansfield, 1986). This report provides additional information about the language tasks that appear on the current TOEFL and is an important reference for an understanding of the content validity of the test. Changes in the test, designed to make it more reflective of communicative competence, are enumerated on pages 92-93 of that report.

Some of the earliest and most basic TOEFL research attempted to match performance on the test with other indicators of English language proficiency, thus providing criterion-related evidence of TOEFL's validity. In some cases these indicators were tests themselves.

A study conducted by Maxwell (1965) at the Berkeley campus of the University of California found a .87 correlation between total scores on TOEFL and the English proficiency test used for the placement of foreign students at that campus. This correlation was based on a total sample of 238 students (202 men and 36 women, 191 graduates and 47 undergraduates) enrolled at the university during the fall of 1964. Upshur (1966) conducted a study to determine the correlation between TOEFL and the Michigan Test of English Language Proficiency. This was based on a total group of 100 students enrolled at San Francisco State College (N = 50), Indiana University (N = 38), and Park College (N = 12) and yielded a correlation of .89. Other studies comparing TOEFL and Michigan Test scores have been done by Pack (1972) and Gershman (1977). In 1966 a study was carried out at the American Language Institute (ALI) at Georgetown University comparing scores on TOEFL with scores on the ALI test developed at Georgetown. The correlation of the two tests for 104 students was .79.

In addition to comparing TOEFL with other tests, some of these studies included investigations of how performance on TOEFL related to teacher ratings. In the ALI Georgetown study the correlation between TOEFL and these ratings for 115 students was .73. Four other institutions reported similar correlations. Table 13 gives the data from these studies. At each of the institutions (designated by code letters in the table) the students were ranked in four, five, or six categories based on their proficiency in English as determined by university tests or other judgments of their ability to pursue regular academic courses (American Language Institute, 1966).

Table 13. Correlations of Total TOEFL Scores with University Ratings

University	Number of Students	Correlations with Ratings
A	215	.78
B	91	.87
C	45	.76
D	279	.79

Looking at criterion-related evidence of TOEFL's validity from a different point of view, two additional studies compared performance on TOEFL with other, more direct indicators of ability to use English. The first of these studies (Pitcher and Ra, 1967) was designed to determine the degree to which TOEFL scores, especially on the Writing Ability section (one of the five sections used prior to 1976), correlated with ratings of actual writing. A group of 310 foreign students enrolled in colleges and universities throughout the country were given all parts of the TOEFL test and then asked to write four essays, one on each of four assigned topics. Every essay was rated twice on a four-point scale of overall quality by each of two readers working independently. The correlation between the sum of the first four ratings given to a student with the sum of the second four ratings was .85, indicating a reader reliability of .92 for the sum of the eight ratings. The correlations of the five TOEFL section scores used at that time with the sum of the eight ratings of writing ability as indicated by the essays ranged from a low of .56 for the Listening Comprehension section to a high of .74 for both the English Structure and Writing Ability sections. These correlations are shown in Table 14.

Table 14. Correlations of TOEFL Scores with the Sum of Theme Ratings (N = 310)

TOEFL Scores	Correlation with Sum of Ratings
Listening Comprehension	.56
English Structure	.74
Vocabulary	.71
Reading Comprehension	.65
Writing Ability	.74
TOEFL Total	.78

The correlation of total scores on TOEFL with the sum of the eight ratings was .78, quite similar to the correlations reported in Table 13, especially those for universities A, C, and D.

In the second of these studies on the five-section test, Pike (1976) investigated the relationship of TOEFL and its subsections to a number of alternate criterion measures, including writing samples, cloze tests, oral interviews, and sentence combining exercises. In general, the results confirmed a close relationship between the five sections of TOEFL and the English skills they were intended to measure. Among the most significant findings of this study were the correlations between TOEFL subscores and two nonobjective measures: oral interviews and writing samples (essays).

Table 15 gives the correlation coefficients for the three language groups participating in the study. Moreover, the figures are shown for both the total interview ratings and the grammar and vocabulary subscores; the essay ratings are listed according to two different scoring schemes—one focusing on essay content and one on essay form. The strong correlations and common variances found in Pike's study between some of the sections of the TOEFL test led to the combining and revising of those sections for the three-part version of the test.

Table 15. Correlations of TOEFL Subscores with Interview and Essay Ratings

		Interview			Essay	
		Gram.	Vocab.	Total	Content	Form
Listening Comprehension	Peru	.84	.84	.84	.83	.91
	Chile	.76	.75	.78	.76	.83
	Japan	.84	.83	.82	.59	.72
English Structure	Peru	.86	.87	.87	.86	.92
	Chile	.88	.87	.87	.88	.98
	Japan	.70	.69	.71	.55	.81
Vocabulary	Peru	.82	.83	.82	.80	.84
	Chile	.77	.77	.75	.74	.83
	Japan	.55	.62	.59	.45	.66
Reading Comprehension	Peru	.88	.87	.87	.84	.85
	Chile	.74	.76	.75	.67	.82
	Japan	.62	.62	.62	.61	.73
Writing Ability	Peru	.86	.85	.86	.85	.93
	Chile	.79	.78	.75	.77	.88
	Japan	.59	.62	.60	.64	.73

Other criterion-related evidence of TOEFL's validity are those studies that have focused on the relationship of TOEFL to some widely used aptitude tests. To some extent, however, the findings of these studies also lend support to the construct-related validity evidence by showing the extent to which the test measures a particular construct or trait—in this case, proficiency in English as a foreign language. One of these studies (Angelis, Swinton, and Cowell, 1979) compared the performance of nonnative speakers of English on TOEFL with their performance on the verbal portions of the GRE Aptitude (now General) Test (graduates) or both the SAT and the Test of Standard Written English (undergraduates). As indicated in Table 16, the GRE verbal performance of the nonnative speakers was extremely low in comparison with the performance of the native speakers. Similar results were reported for undergraduates (Table 17).

Table 16. GRE Verbal Score Comparisons

	Mean	S.D.	Rel.	S.E.M.
Experimental Group (Nonnatives) (N = 186)	274	67	.78	30
Native Speakers (N = 1,495)	514	128	.94	32

Table 17. SAT and TSWE Score Comparisons

	SAT Verbal			
	Mean	S.D.	Rel.	S.E.M.
Experimental Group (Nonnatives) (N = 210)	269	67	.77	33
Native Speakers (N = 1,765)	425	106	.91	32
	TSWE			
	Mean	S.D.	Rel.	S.E.M.
Experimental Group (Nonnatives) (N = 210)	28	8.8	.84	4
Native Speakers (N = 1,765)	42.35	11.09	.89	3.7

Wilson (1982) conducted a similar study of all GRE, TOEFL, and GMAT examinees during a two-year period extending from 1977 to 1979. These results, depicted in Table 18, combined with those obtained in the earlier study by Angelis, Swinton, and Cowell, warrant an important conclusion for admissions officers: Verbal aptitude test scores of nonnative examinees are significantly lower than the scores earned by native English speakers. On the other hand, quantitative aptitude scores are not greatly affected by a lack of language proficiency. Further, analyses of each study show that only when TOEFL scores reach approximately the 625 level do verbal aptitude test scores of foreign candidates reach the level normally obtained by native English speakers.

Table 18. GRE and GMAT Score Comparisons, 1977-79

Test Population	N	Verbal Mean	Quantitative Mean	Analytical Mean
GRE all	831,650	479	518	496
GRE foreign ESL (TOEFL mean = 552)	2,442	345	606	400
GMAT all	563,849	26	27	NA
GMAT foreign ESL (TOEFL Mean = 542)	3,918	15.7	29	NA

To provide guidelines for those who may be evaluating applicants presenting scores from more than one of the above tests, Angelis, Swinton, and Cowell (1979) conducted special analyses. Results indicated that, for graduate-level applicants, 475 on TOEFL is a critical deciding point for interpretations of GRE verbal scores. Applicants above that level tend to have GRE verbal scores that, although lower than scores for natives, fall within a range for which estimates of verbal ability can be made for students with similar TOEFL scores. Those below the 475 TOEFL level tend to have such low GRE verbal scores that such estimates cannot easily be made. At the undergraduate level, 435 on TOEFL is a key decision point. SAT verbal scores for applicants below that level are not likely to be informative. Similarly, Powers (1980) found that a TOEFL score of 450 is required before GMAT verbal scores begin to discriminate among examinees.

All of these TOEFL score levels are below the ranges normally considered indicators of proficiency required for academic study (see page 19), thus demonstrating that valid comparisons of verbal aptitude can be made for acceptable applicants so long as the applicants compared have similar TOEFL scores.

As noted earlier, interpreting the relationship between language proficiency and aptitude and achievement test scores in verbal areas can be complex. Few of even the most qualified foreign applicants approach native proficiency in English. Thus, verbal aptitude scores of nonnative English speakers are likely to be depressed even when TOEFL test scores are high. Only when TOEFL scores are at the native speaker level (625 or above) does the distribution of scores on a verbal aptitude test become similar to the distribution obtained by native English speakers. Cultural factors and cross-national differences in educational programs may also affect performance on tests of verbal ability.

The TOEFL program has published three research reports that can assist in evaluating the effect of language proficiency on an applicant's performance on specific standardized

tests. *The Performance of Nonnative Speakers of English on TOEFL and Verbal Aptitude Tests* (Angelis, Swinton, and Cowell, 1979) gives comparative data about the performance of a group of foreign students on TOEFL and either the GRE verbal or the SAT verbal and the Test of Standard Written English. *The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language* (Powers, 1980) compares performance on TOEFL and GMAT. Additional information and recent comparisons are available in *GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL* (Wilson, 1982).

TOEFL is a three-part test. Support for the three-part division is evidenced by the pattern of correlations between each of the TOEFL sections and other tests (Angelis, Swinton, and Cowell, 1979). The GRE verbal score correlates highest with the Vocabulary and Reading Comprehension section of TOEFL (.623). The same section correlates highest (.681) with the SAT verbal score. This is to be expected since both verbal aptitude tests rely heavily on reading and vocabulary. For the College Board's Test of Standard Written English, the highest correlation (.708) is with Section 2 of TOEFL, Structure and Written Expression. Again, this is to be expected since the Test of Standard Written English uses knowledge of grammar and related linguistic elements as indicators of writing ability. In all three cases, the lowest correlations are those with Section 1 of TOEFL, Listening Comprehension. Since none of the other tests includes items that attempt to measure ability to understand spoken English, this again is to be expected.

In another study, comparing performance of nonnative speakers of English on TOEFL and the Graduate Management Admission Test, Powers (1980) reports the same pattern of correlations. As indicated in Table 19, the highest GMAT verbal-TOEFL correlation is that for the Vocabulary and Reading Comprehension section. Correlations for Section 2 are slightly lower and those for Section 1 (listening) are the lowest. The fact that the correlations for the quantitative section of the GMAT are the lowest of all (ranging from .29 to .39) provides support for the discriminant validity of TOEFL as a measure of verbal skills in contrast to quantitative skills.

A final conclusion of this study, which touches on the overall relationship between the two tests, is that a minimum TOEFL total score of approximately 450 is required before GMAT verbal scores begin to discriminate among candidates with respect to verbal ability as measured by the GMAT.

Table 19. Simple Correlations Between GMAT and TOEFL Scores

	TOEFL Scores			
GMAT Score	Listening Comprehension	Structure and Written Expression	Vocabulary and Reading Comprehension	Total
GMAT Verbal	.58	.66	.69	.71
GMAT Quantitative	.29	.37	.39	.39
GMAT Total	.52	.61	.64	.66

In early attempts to obtain construct-related evidence of validity for TOEFL, two studies were conducted comparing the performance of native and nonnative speakers of English on the test. Angoff and Sharon (1970) found that the mean scores on TOEFL of native speakers in the United States were much higher than those of foreign students who had taken the same test. Evidence that the test was quite easy for the American students is found in the observations that their mean scores were not only high but homogeneously high relative to those of the foreign students; that their score distributions were highly negatively skewed; and that a high proportion of them earned maximum or near-maximum scores on the test.

A more detailed study of native speaker performance on TOEFL was conducted by Clark (1977). Once again, performance on the test as a whole proved similar to that of the native speakers included in the Angoff and Sharon study. The mean raw score for the native speakers, who took two different forms of TOEFL, was 134 (out of 150). This compared to mean scores of 88 and 89 for the nonnative speakers who had originally taken the same forms. However, additional analysis showed that the native speakers did not perform equally well on all three sections of the test. While their mean percentage of incorrect items for the Listening Comprehension section of the test was 4.4 percent, the percentages for the Structure and Written Expression and Vocabulary and Reading Comprehension sections were 14.6 and 12.1, respectively.

Such information is useful for test development since it provides guidelines on which to base evaluations of questions at the review stage. A detailed analysis of the relevant questions from the Clark study indicated that, even though not all the native speakers chose the correct answers, the points tested represent basic features of English that would appear to be indispensable for effective academic work at the undergraduate level.

Other TOEFL Programs

Test of Written English (TWE®)

This thirty-minute essay test provides the examinee with an opportunity to perform academic writing tasks similar to those required of students in North American universities. This includes the ability to generate and organize ideas on paper, to support those ideas with examples or evidence, and to use the conventions of standard written English.

The examinee is given one topic on which to write. As with other TOEFL test items, the TWE essay questions are developed by specialists in English or ESL, and each essay question is field-tested and reviewed by a committee of composition specialists, the TWE Core Reader Group. A pretested topic will be approved for use in the TWE only if it elicits a range of responses at a variety of proficiency levels,

does not advantage or disadvantage any examinee or group of examinees, and does not require special subject matter knowledge. The essay questions are also reviewed for racial and cultural bias and content appropriateness according to established ETS sensitivity review procedures.

The essays are read by trained English and ESL composition specialists who assign scores based on a six-point, criterion-referenced scoring guide. Each essay is read by at least two readers working independently. The score an examinee receives is the average of the scores assigned to the essay by the readers.

At this time the Test of Written English score is not incorporated into the total TOEFL score. Instead, a separate TWE score is reported on the TOEFL score report that is mailed to the examinee and to the institution(s) to which the examinee requests that his or her scores be sent. Designated score recipients receive a copy of the TWE Scoring Guide, which describes the proficiency levels associated with the six holistic score points. Sample essays at the six score levels are published in the *TOEFL Test of Written English Guide*.

TWE test results can assist institutions in evaluating the academic writing proficiency of their ESL and EFL students and in placing these students in appropriate writing courses.

Additional information and copies of the TWE *Guide* may be obtained by writing to the TOEFL Program Office, P.O. Box 6155, Princeton, NJ 08541-6155.

Test of Spoken English (TSE®)

The Test of Spoken English was developed by ETS under the direction of the TOEFL Policy Council to provide a reliable measure of proficiency in spoken English. The TSE test affords educational institutions, government agencies, and other organizations a means of predicting the speaking ability of nonnative speakers of English for whom oral communication can be critical to performance in the classroom or in other professional situations.

TSE is administered twelve times a year on the same dates as the TOEFL test. The test takes approximately thirty minutes and can be administered to individuals with cassette or reel-to-reel tape recorders or to a group using a language laboratory or portable language laboratory equipment provided by the TSE program.

The TSE test requires examinees to demonstrate their spoken English proficiency by responding orally under timed conditions to a variety of printed and aural stimuli that are designed to elicit both controlled and spontaneous reactions. All examinee responses are recorded on tape.

The test has seven sections, each involving a particular speech activity. The first section is an unscored "warm-up" in which the examinee responds to brief general questions. In the second section, the examinee reads aloud a printed passage of approximately 125 words and is told that scoring will be based on pronunciation and overall clarity of speech. Time is allowed for preliminary silent reading of the passage.

In the third section, the examinee sees a series of ten partial sentences and is asked to complete each sentence orally in a way that conveys meaning and is grammatically correct. The fourth section consists of six photographs or drawings that "tell a continuous story." After studying the drawings briefly, the examinee is asked to tell the story that the pictures show and to include as much detail as possible. In section 5, the examinee answers a series of spoken questions about the content of a single picture.

Section 6 consists of a series of spoken questions intended to elicit free, and somewhat lengthy, responses from the examinee. The questions require both descriptions of common objects and open-ended expressions of opinion on familiar issues. The linguistic quality and adequacy of communication, rather than specific knowledge revealed, are considered in scoring this section. In the seventh and final section, the examinee sees a printed schedule, such as for an imaginary course or conference, and is asked to describe the program (e.g., meeting time, location, subject) as though speaking to a group of students or colleagues.

The TSE answer tapes are rated by trained specialists; depending on the section, responses to each test section are given ratings for two or more of the following categories: pronunciation, grammar, fluency, and overall comprehensibility. Separate scores for these categories are computed from the ratings, and official score reports are sent to institutions designated by the examinees.

The TSE test has broad applicability in situations where it is important to have an accurate indication of an individual's level of speaking proficiency. For example, academic institutions may use TSE scores to evaluate the spoken English of applicants for teaching assistantships or other academic positions; professional licensing agencies may use the scores for selection and certification purposes in certain health-related professions. Because subscores are provided, TSE may be used to diagnose areas of strength or weakness in spoken English. The test is not targeted to a single academic discipline, field of employment, or other specialized language usage; the examinee, regardless of native language or specialization, is required to demonstrate a level of general speaking proficiency.

Additional information about interpretation and use of TSE scores may be obtained from the *TSE Manual for Score Users*, which is available from the TSE Program Office, P.O. Box 6157, Princeton, NJ 08541-6157, USA.

Speaking Proficiency English Assessment Kit (SPEAK®)

The Speaking Proficiency English Assessment Kit (SPEAK) was developed by the TOEFL program to provide—to institutions wishing to administer and score the TSE test themselves—a valid and reliable instrument for assessing the English speaking proficiency of people who are not native speakers of the language. The kit includes SPEAK test form 1, testing materials for thirty examinees, a description of procedures for administering the test, and materials for training SPEAK raters. The training materials include eight rater-training cassettes and six rater-testing cassettes, as well as complete instructions for the training program in the *Guide to Speak*. Raters determine whether they have mastered the necessary rating skills by comparing the ratings they assign to the rater-testing cassettes with the correct ratings provided in the *Guide*.

SPEAK test results can be used to evaluate the speaking proficiency of nonnative English-speaking applicants for teach-

ing assistantships, to measure improvement in speaking proficiency over a period of time, or to identify teaching assistants and others who need additional instruction in English.

SPEAK test forms 2, 3, and 4, different editions of the test, are also available to purchasers of the SPEAK kit. The additional forms make it possible to reevaluate the spoken English of individuals who have been tested earlier with the original SPEAK test. SPEAK test forms 2, 3, and 4 are well suited for monitoring the performance of students in English language programs and for retesting individuals who, for example, did not meet the oral proficiency requirements for positions as university teaching assistants on an earlier test. SPEAK test forms 2, 3, and 4 are each packaged with thirty test books, thirty *Examinee Handbooks*, one reel-to-reel tape and one cassette, and fifty rating sheets.

Further information about SPEAK is available from the TOEFL Program Office, P.O. Box 6155, Princeton, NJ 08541-6155, USA.

Secondary Level English Proficiency (SLEP®) Test

The Secondary Level English Proficiency test is designed for students entering grades seven through twelve who are nonnative speakers of English. The test is a measure of proficiency in two primary areas: understanding spoken English and understanding written English. It is based on the assumption that language proficiency is a critical factor in determining the degree to which students can benefit from instruction. SLEP is not an aptitude test or a measure of academic achievement; it is a measure of English language ability. The results of the test can be very helpful in making placement decisions related to assignment to ESL classes, placement in a mainstream English-medium program, or exit from an ESL program. Because the SLEP scale is sensitive to small gains in language skills, the test can be useful for program evaluation purposes.

There are three different forms of the SLEP test, all made to the same test specifications, equated, and norm referenced. Each test form contains 150 multiple-choice questions of eight different types and is divided into two sections, listening comprehension and reading comprehension. The questions in the first section of the test use taped samples of spoken English to test listening comprehension and do not rely heavily on written materials. The questions in the second section measure vocabulary, grammar, and overall reading comprehension and are based on written and visual materials. Answer sheets are easily scored, and technical data for interpreting test results are provided in the *SLEP Test Manual*.

SLEP testing materials are available for direct purchase. The basic package of testing materials for each form contains twenty SLEP test books, one hundred three-ply answer sheets, a copy of the *SLEP Test Manual*, and a cassette recording of the listening comprehension section questions. Each item in the basic package may also be purchased separately.

Additional information and copies of SLEP publications may be obtained by writing to SLEP Program Office, P.O. Box 6155, Princeton, NJ 08541-6155, USA.

Procedures at Test Centers

Standard, uniform procedures are important in any testing program, but are essential for an examination that is given worldwide. Therefore, the TOEFL program provides detailed guidelines for test center supervisors to ensure uniform administrations. *Preparing for a TOEFL/TSE Administration* is mailed to test supervisors well in advance of the test date. This publication describes the arrangements the supervisor must make to prepare for the test administration, including selecting testing rooms and the associate supervisors and proctors who will be needed on the day of the test.

The *Manual for Administering TOEFL*, included with every shipment of test materials, describes appropriate seating plans, the kind of equipment that should be used for the Listening Comprehension section, the priorities for admitting examinees to the testing room, and instructions for distributing and collecting test materials. It also contains detailed instructions for the actual administration of the test.

The TOEFL program office works with test center supervisors to ensure that the same practices and procedures are followed at all centers, and it conducts workshops during which supervisors can discuss with TOEFL staff the procedures for administering the test. TOEFL staff respond to all inquiries from supervisors and examinees regarding circumstances or conditions associated with test administrations, and they investigate all complaints received about specific administrations.

Measures to Protect Test Security

To protect the validity of the test scores, the TOEFL office constantly reviews and refines procedures to increase the security of the test before, during, and after its administration.

Because of the importance of TOEFL test scores to examinees and to institutions, there are inevitably some individuals who will engage in practices designed to increase their reported scores. The careful selection of supervisors, a high proctor-to-examinee ratio, and carefully developed procedures for the administration of the test (explained in the *Manual for Administering TOEFL)* are measures designed to prevent or discourage examinee attempts at impersonation, copying, theft of test materials, and the like, and thus to protect the integrity of the test for all examinees and score recipients.

Identification Requirements

Strict admission procedures are followed at all test centers to prevent attempts by some examinees to have others with greater proficiency in English impersonate them at a TOEFL administration. To be admitted to a test center, every examinee must present an official document with a recognizable photograph and a completed photo file record with a recent photo attached. Although the passport is the basic document that is acceptable at all test centers, other *specific* photobearing documents are acceptable for individuals who may not be expected to have passports or who are taking the test in their own countries.

Through foreign embassies in the United States and TOEFL supervisors in foreign countries, the TOEFL office continually verifies the names of official, secure, photobearing identification documents used in each country, such as national identity cards, registration certificates, and work permits. In the Friday and Saturday Testing Programs, each admission ticket contains a statement specifying the documents that will be accepted at TOEFL test centers in the country in which the examinee is registered to take the test. This information is computer-printed on a red field to ensure that it will be seen. (The same information is printed on the attendance roster prepared for each center.) Following is a sample of the statement that appears on admission tickets for Venezuela.

> YOUR PASSPORT. CITIZENS OF VENEZUELA MAY USE NATIONAL IDENTITY CARD OR LETTER AS DESCRIBED IN THE BULLETIN.

Complete information about identification requirements is included in all editions of the *Bulletin of Information for TOEFL and TSE*. (The United States/Canada edition also notes that officials at selected test centers in these countries may photograph the examinees on the day of the test.)

Photo File Records. The TOEFL program has implemented additional procedures designed to ensure the validity of TOEFL test scores by attempting to eliminate the possibility of impersonation at test centers. These procedures require each examinee to present a completed photo file record to the test center supervisor before being admitted to the testing room. The photo file record contains the examinee's name, registration number, test center code, and signature, as well as a recent photo that clearly identifies the examinee (that is, the photo must look exactly like the examinee, with the same hairstyle, with or without a beard, and so forth). The photo file records are collected at the test center and returned to ETS. Institutions and agencies that notice an inconsistency between test scores and English performance may request copies of the TOEFL photo file records to verify identity (see "Examinee Identification Service," page 35).

Checking Names and Signatures

To prevent examinee attempts to exchange answer sheets or to grid another person's name (for whom he or she is taking the test) on the answer sheet, supervisors are asked to check the gridding of names on the answer sheets and also to check signatures on the answer sheets with those on the identification documents before examinees leave the testing room.

Supervision of Examinees

Supervisors and proctors are instructed to exercise extreme vigilance during a test administration to prevent examinees from giving or receiving assistance in any way.

In addition, the *Manual for Administering TOEFL* advises supervisors about assigning seats to examinees. To prevent copying from notes or other aids, examinees may not have anything on their desks but their test books, answer sheets, pencils, and erasers. They are not permitted to make notes or marks of any kind in their test books. ("Dismissal Notice" forms are used to report examinees who violate procedures. An examinee signs the notice to document the violation and to indicate that he or she understands that a violation of procedures has occurred and that the answer sheet may not be scored.)

If a supervisor is certain that someone has given or received assistance, the supervisor has the authority to dismiss the examinee from the testing room; scores for dismissed examinees will not be reported. If a supervisor **suspects** someone of cheating, the examinee is warned about the violation, is asked to sign a "Dismissal Notice," and must move to another seat selected by the supervisor. A description of the incident is written on the Supervisor's Irregularity Report, which is returned to ETS with the answer sheet. Both suspected and confirmed cases of cheating are investigated by the Test Security Office at ETS. (See "Scores of Questionable Validity," page 13.)

Turning back to another section of the test, working on a section in advance, or continuing to work on a section after time is called are not permitted and are considered cheating. (To assist the supervisor, a large number identifying the section being worked on is printed at the top of each page of the test book.) Supervisors are instructed to warn anyone found working on the wrong section and to ask the examinee to sign a "Dismissal Notice."

Preventing Access to Test Materials

To ensure that examinees have not seen the test material in advance, a new form of the test is developed for each Friday and Saturday Testing Program administration.

To prevent the theft of test materials, procedures have been devised for the distribution and handling of these materials. Test books, answer sheets, and Listening Comprehension recordings are sent to test centers in sealed boxes and are placed in secure, locked storage that is inaccessible to unauthorized persons. Test books are individually sealed and are packed in sealed plastic bags. Supervisors are directed to count the test books several times—upon receipt, during the test administration, and after the test is finished. No one is permitted to leave the testing room until all materials have been accounted for. Except for "disclosed" administrations, when examinees may obtain the test book (see "Test Forms Available to Examinees," page 35), supervisors must follow detailed directions for returning the test materials. Materials are counted upon receipt at ETS, and the Test Security Office investigates all cases of missing test materials.

TOEFL Publications and Services

Information about ordering the publications and services described below may be obtained by writing to TOEFL, P.O. Box 6155, Princeton, NJ 08541-6155, USA.

Friday and Saturday Testing Programs

Bulletin of Information for TOEFL and TSE

This free publication is the primary source of information for individuals who wish to take the TOEFL and/or TSE test(s) at Friday or Saturday Testing Program administrations. The *Bulletin* tells examinees how to register for the tests and lists the test centers for both programs, provides a brief description of the tests, and explains score reporting and other procedures. It also contains the TOEFL/TSE calendar, which includes the test dates, registration deadline dates, and mailing dates for official score reports. In addition, there are practice questions, detailed instructions for filling out the answer sheet on the day of the test, an explanation of procedures to be followed at the test center, and information about interpreting scores on the tests.

The *Bulletin* also includes information about the Test of Written English (TWE).

There is currently (1990) a single edition for the United States and Canada, and there are several editions of the Overseas *Bulletin* (one general and others specific to countries or regions). The bulletins for specific countries or regions are prepared to facilitate registration through the program's official representatives for those countries. Because of the variations in arrangements with TOEFL/TSE representatives for the individual countries or regions and government regulations in those countries, examinees must obtain the appropriate edition of the *Bulletin* for the location of the test center at which they wish to test. Each edition of the *Bulletin* contains directions for obtaining the appropriate editions for all countries.

Copies of the Overseas *Bulletin* are available at many overseas counseling centers, United States embassies, and offices of the United States Information Service. In countries and regions where registration is handled by TOEFL representatives, the representatives distribute appropriate editions of the Overseas *Bulletin* to examinees and local institutions.

Examinee's Order Form for Study Materials and the Bulletin of Information

Most institutions that require TOEFL and/or TSE scores find it convenient to send a copy of the Examinee's Order Form to their foreign applicants. The form gives instructions for obtaining the appropriate edition of the *Bulletin* and includes the TOEFL and TSE test date calendars. It also provides information about ordering official TOEFL study materials.

Test Center Reference List

The test center reference list provides TOEFL and TSE test dates, registration deadline dates, score report mailing dates, and test center locations for the Friday and Saturday Testing Programs. It also tells how to obtain the appropriate edition of the *Bulletin*. The free list is distributed at the beginning of each testing year to institutions and organizations that use TOEFL and TSE test scores.

Test and Score Manual

The *Manual* is revised every two years. The edition current at the time a score report is released is indicated on the score report. Copies of each new edition of the *Manual* are sent to institutions receiving TOEFL test scores; additional copies are available upon request at no charge. *Guidelines for the Use of TOEFL Scores* is a booklet containing critical information from the *Manual* about interpreting TOEFL test results. It is a free publication.

Institutional Testing Program

A brochure describing this program is distributed each year to interested institutions worldwide. The brochure contains a summary of the TOEFL and Pre-TOEFL services available, a description of the test, details about ETS policy regarding testing and the release of examinee score data, and an order form.

With each order, copies of the *Examinee Handbook*, testing materials, and directions for administering the test are sent in quantities sufficient for the number of individuals to be tested.

TOEFL Study Materials

The study materials described here are official publications of the TOEFL program. They were prepared by the TOEFL program staff to help students understand the specific linguistic skills the TOEFL test measures and become familiar with the multiple-choice formats used.

Sample Test

The *Sample Test* is about half the length of the actual test and contains the same kinds of questions. It consists of:

- a test book with seventy-five practice questions
- an answer sheet and a list of the correct answers
- a 33-1/3 rpm plastic recording of the Listening Comprehension questions

Test Kits

Understanding TOEFL includes an actual form of the test and explanatory materials. The test kit contains:

- a test book with 150 questions
- an answer sheet
- a cassette recording of the Listening Comprehension section and an analysis of the Listening Comprehension questions
- a workbook with a general introduction to the test, a description of each section in the test, discussions of the

different types of questions, a list of the correct answers, and an explanation of every question. It also gives complete instructions for taking the test.

Listening to TOEFL focuses on Section 1, the Listening Comprehension section of the test. The test kit contains:

- 175 practice questions
- two complete Listening Comprehension sections from actual TOEFL tests
- two full-length TOEFL tests
- three sixty-minute cassettes
- a workbook giving instructions for using the cassettes, studying the practice questions, and taking the listening tests and complete tests. Scripts for the practice questions and the two listening tests, answer sheets, and lists of the correct answers are provided.

Reading for TOEFL focuses on Section 3, the Vocabulary and Reading Comprehension section of the test. The test kit contains:

- Vocabulary and Reading Comprehension questions from four TOEFL tests
- two full-length TOEFL tests
- one sixty-minute cassette
- a workbook giving instructions for studying the Vocabulary and Reading Comprehension questions and for taking the complete tests. Answer sheets and lists of correct answers are also included.

A companion classroom pack for each test kit is available as an instructional tool for ESL and language laboratory use, as well as for workshops. Containing ten test books and/or workbooks, the classroom packs are an economical way to familiarize students with the TOEFL test. Instructors need to purchase only one copy of a test kit (containing recorded material) for use with the companion classroom pack.

Test Forms Available to Examinees

Examinees who actually take the test on specific dates announced in advance by the TOEFL office may obtain the test books used at these administrations free of charge. In addition, these examinees may order a list of the correct answers, a cassette recording of Section 1 (Listening Comprehension), and a copy of their answer sheet with the raw scores marked.

An order form with information about how to order and pay for these materials is printed on the inside back covers of the test books for these test administrations. The availability of this service is subject to change without notice.

Fee Voucher Service

The TOEFL program offers the fee voucher service for the convenience of those that pay TOEFL and/or TSE test fees for some or all of their students or applicants. Upon receipt of a completed Request Form for Fee Voucher Service, cards in the quantity ordered by the institution or agency are sent to the address given on the form. Each fee voucher card shows the name and code number of the participating institution and is valid only for the specified testing year and the specific program (TOEFL or TSE) specified thereon. It is not an acceptable document for admission to the test center. The institution receives an itemized invoice following each test administration for the number of voucher cards it issued and submitted in lieu of personal payment by the persons taking the test(s).

Computer Tape Score-Reporting Service

A computer tape score-reporting service for TOEFL official score-report recipients is available by subscription. The service provides TOEFL score reports on magnetic tapes to participating institutions and agencies for a nominal annual charge. Because the TOEFL office releases score reports only upon the written request of the examinees, the tape prepared for each institution or agency contains only the score records of TOEFL examinees who requested that their scores be reported to that institution or agency.

The score reports are in single record format on 9 track/1600 b/p/i magnetic tapes, without labels, and contain all the information printed on the official score reports for each examinee who designated the institution as a score recipient. Each tape contains the score records requested during two-month processing periods.

Subscription to the computer tape score-reporting service is for one year (July to June) and may begin at any time during the year (however, only tapes of upcoming processing periods can be mailed). Further information and subscription forms may be obtained from the TOEFL office.

Examinee Identification Service

The photo file record is collected by the test center supervisor from each examinee before he or she is admitted to the testing room.

The examinee identification service provides photo identification of examinees taking the TOEFL test. If there is reason to suspect an inconsistency between a high test score and relatively weak English proficiency, an institution or agency that has received either an official score report from ETS or an examinee's score record from an examinee may request a copy of that examinee's photo file record up to eighteen months following the test date shown on the score report. The written request for examinee identification must be accompanied by a photocopy of each examinee's score record or official score report.

Program Support for External Studies

The TOEFL program will make available certain types of test data or perform analyses of pertinent data requested by external researchers for studies relating to language proficiency. The researchers must agree to protect the confidentiality of the data, agree to assume responsibility for the analyses and conclusions of the studies, and agree to reimburse the TOEFL program for the costs associated with the compilation and formatting of the data.

TOEFL program funding of independent research, if requested and granted, is usually limited to providing test materials and related services without charge and/or the cost of the data access and data analysis.

Individuals interested in utilizing TOEFL test data for research studies should write to the TOEFL program office.

TOEFL Research Reports

The Performance of Native Speakers of English on the Test of English as a Foreign Language. John L. D. Clark. November 1977. Report 1. Discusses the results of the administration of 1977 forms of TOEFL to native speakers of English just prior to their graduation from a college-preparatory high school program; reinforces earlier findings that TOEFL is not psychometrically appropriate for native speakers of English.

An Evaluation of Alternative Item Formats for Testing English as a Foreign Language. Lewis W. Pike. June 1979. Report 2. Describes an extensive research study conducted from 1972 to 1974 that was designed to explore possible changes in the format and content of TOEFL; contributed to the restructuring of TOEFL beginning in 1976.

The Performance of Non-Native Speakers of English on TOEFL and Verbal Aptitude Tests. Paul J. Angelis, Spencer S. Swinton, and William R. Cowell. October 1979. Report 3. Gives the results of a study in which 400 graduate and undergraduate applicants took TOEFL and either the GRE verbal or the SAT verbal and the Test of Standard Written English; includes comparative data on performance across tests.

An Exploration of Speaking Proficiency Measures in the TOEFL Context, John L. D. Clark and Spencer S. Swinton. October 1979. Report 4. Describes a three-year study involving the development and experimental administration of test formats and item types aimed at measuring the English speaking proficiency of nonnative speakers; results grouped into a prototype "Test of Spoken English."

The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language. Donald E. Powers. December 1980. Report 5. Analyzes the performance of 6,000 nonnative speakers of English on TOEFL and GMAT; provides support of the basic differences in the two tests and indicates expected GMAT scores for examinees with differing levels of English language proficiency.

Factor Analysis of the Test of English as a Foreign Language for Several Language Groups. Donald E. Powers and Spencer S. Swinton. December 1980. Report 6. Provides evidence that three major factors underlie performance on TOEFL; suggests these factors may be interpreted differently for several language groups.

The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings. John L. D. Clark and Spencer S. Swinton. December 1980. Report 7. Examines the performance of teaching assistants on the Test of Spoken English in relation to their classroom performance as judged by students; reports that TSE is a valid predictor of oral language proficiency for nonnative English-speaking graduate teaching assistants.

Effects of Item Disclosure on TOEFL Performance. Paul J. Angelis, Gordon A. Hale, and Lawrence A. Thibodeau. December 1980. Report 8. Assesses the effects of test disclosure by examining the performance on TOEFL when a subset of items have been studied prior to an administration; provides separate results by language group and by item type.

Item Performances Across Native Language Groups on the Test of English as a Foreign Language. Donald L. Alderman and Paul W. Holland. August 1981. Report 9. Examines the performance of different native language groups on TOEFL items; discusses implications for the interpretation and examination of item performance by groups.

Language Proficiency as a Moderator Variable in Testing Academic Aptitude. Donald L. Alderman. November 1981. Report 10. Demonstrates the role of language proficiency as a moderator variable in assessing academic aptitude; a moderately strong correlation develops between verbal aptitude tests in the native and second languages when TOEFL scores indicate high second language proficiency.

A Comparative Analysis of TOEFL Examinee Characteristics, 1977-1979. Kenneth M. Wilson. October 1982. Report 11. Provides detailed comparative information about the personal characteristics, academic aspirations, and test scores of TOEFL examinees by region, native country, and native language.

GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL. Kenneth M. Wilson. October 1982. Report 12. Analyzes the performance of examinees taking TOEFL and either the GRE Aptitude Test or GMAT; provides further documentation of the relationship between English language proficiency and aptitude test scores earned by foreign students.

The Test of Spoken English as a Measure of Communicative Proficiency in the Health Professions. Donald E. Powers and Charles W. Stansfield. November 1982. Report 13. Provides results of using a set of procedures for determining standards of language proficiency in testing pharmacists, physicians, veterinarians, and nurses for validating the use of TSE in health-related professions.

A Manual for Assessing Language Growth in Instructional Settings. Spencer S. Swinton. November 1982. Report 14. Describes a methodology for determining the true gains in proficiency that can be expected for students who enter English language training programs at different TOEFL score levels; discusses how the relationship between gains and time enrolled in a program can be used to advise students.

A Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students. Brent Bridgeman and Sybil Carlson. September 1983. Report 15. Describes a survey of faculty in 190 departments at 34 U.S. and Canadian universities with high foreign student enrollments; respon-

dents indicated a desire to use scores on a direct writing sample to supplement admissions and placement decisions.

Summaries of TOEFL Studies, 1963-1982. Gordon A. Hale, Charles W. Stansfield, and Richard P. Duran. November 1983. Report 16. Includes approximately eighty summaries of empirical research studies involving TOEFL, as well as descriptive papers that provide a perspective on the history and development of the test.

TOEFL from a Communicative Viewpoint on Language Proficiency: A Working Paper. Richard P. Duran, Michael Canale, Joyce Penfield, Charles W. Stansfield, and Judith E. Liskin-Gasparro. February 1985. Report 17. Examines the content characteristics of the TOEFL test from a communicative perspective based on current research in applied linguistics and language proficiency assessment.

A Preliminary Study of Raters for the Test of Spoken English. Isaac I. Bejar. February 1985. Report 18. Examines the scoring patterns of different TSE raters in an effort to develop a method for predicting disagreements; reports that the raters varied in the severity of their ratings but agreed substantially on the ordering of examinees.

Relationship of Admissions Test Scores to Writing Performance of Native and Nonnative Speakers of English. Sybil B. Carlson, Brent Bridgeman, Roberta Camp, and Janet Waanders. August 1985. Report 19. Investigates the relationship between essay writing skills and scores on TOEFL and the GRE General Test obtained from applicants to U.S. institutions.

A Survey of Academic Demands Related to Listening Skills. Donald E. Powers. December 1985. Report 20. Reports the findings of a survey of faculty perceptions of the importance of various listening problems of nonnative English-speaking students.

Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference. Charles W. Stansfield. April 1986. Report 21. Includes invited papers and summaries of the discussions that took place at a conference devoted to the TOEFL program's testing of communicative competence.

Patterns of Test Taking and Score Change for Examinees Who Repeat the Test of English as a Foreign Language. Kenneth M. Wilson. January 1987. Report 22. Analyzes patterns of test taking and score change for examinees who repeated the TOEFL test within twenty-four to sixty months after they first took the test; shows that repeaters registered substantial average net gains in performance, and differences were noted among national-linguistic groups.

Development of Cloze-Elide Tests of English as a Second Language. Winton H. Manning. April 1987. Report 23. Reports on a study to investigate the validity of cloze-elide tests of English proficiency for students similar to the TOEFL candidate population; suggests that cloze-elide tests are good, indirect measures of English language proficiency, comparing very favorably with more commonly used testing procedures.

A Study of the Effects of Item Option Rearrangement on the Listening Comprehension Section of the Test of English as a Foreign Language. Marna L. Golub-Smith. August 1987. Report 24. Provides evidence that scrambling a test question's answer choices produces differences in both the estimated response functions and equating functions.

The Interaction of Student Major-Field Group and Test Content in TOEFL Reading Comprehension. Gordon A. Hale. January 1988. Report 25. Examines the interaction of a student's major-field group with the text content in determining performance on TOEFL reading passages.

Multiple-Choice Cloze Items and the Test of English as a Foreign Language. Gordon A. Hale, Charles W. Stansfield, Donald A. Rock, Marilyn M. Hicks, Frances A. Butler, and John W. Oller, Jr. March 1988. Report 26. Reports on a study to investigate whether multiple-choice (MC) cloze items' involvement of reading comprehension was primary or secondary and the degree to which the items also tapped knowledge of grammar or vocabulary; suggests that skills associated with grammar, vocabulary, and reading comprehension are highly interrelated, as assessed by the TOEFL and the MC cloze items.

Native Language, English Proficiency, and the Structure of the Test of English as a Foreign Language. Philip K. Oltman, Lawrence J. Stricker, and Thomas Barrows. July 1988. Report 27. Assesses the interrelations among TOEFL items for groups of examinees varying in native language and level of English proficiency; concludes that TOEFL construct validity is supported; the test's interpretation varies with examinees' English proficiency, easy and difficult items differ in their potential for diagnosis and global screening, and the dimensionality of the TOEFL test and of competence in English depends on examinees' English proficiency.

Latent Structure Analysis of the Test of English as a Foreign Language. Robert F. Boldt. Report 28. Uses item-response-theory (IRT)-based methods for TOEFL equating; reports a single factor (group) gave a very accurate accounting for the proportions of joint item success.

Context Bias in the Test of English as a Foreign Language. William H. Angoff. January 1989. Report 29. Uses a Mantel-Haenszel analysis to test the hypothesis that TOEFL examinees tested in their native countries are disfavored because of American references in the test; concludes that TOEFL does not disfavor foreign-tested examinees.

Accounting for Random Responding at the End of the Test in Assessing Speededness on the Test of English as a Foreign Language. Charles Secolsky. January 1989. Report 30. Uses two exploratory approaches to determine if TOEFL is speeded according to established criteria; suggests that Section 3 pretest administrations may be slightly speeded but that further confirmation is needed because of the exploratory nature of the methods.

The TOEFL Computerized Placement Test: Adaptive Conventional Measurement. Marilyn M. Hicks. February 1989. Report 31. Reports on the development of an experimental TOEFL computerized placement test using conventional scoring methods based on a testing algorithm that routed examinees through item blocks or testlets and permitted backtracking to review answers and change them.

Confirmatory Factor Analysis of the Test of English as a Foreign Language. Gordon A. Hale, Donald A. Rock, and Thomas Jirele. December 1989. Report 32. Uses confirmatory factor analysis for several language groups; analyses yielded similar results for both domestic and overseas populations. In all cases data supported a two-factor interpretation.

Further information about the TOEFL research reports or about ordering copies of the reports may be obtained by writing to TOEFL, P.O. Box 6155, Princeton, NJ 08541-6155, USA.

References

Alderman, D. L. *TOEFL item performance across seven language groups* (TOEFL Research Report 9). Princeton, NJ: Educational Testing Service, 1981.

American Association of Collegiate Registrars and Academic Officers. *AACRAO-AID participant selections and placement study*. Report to the Office of International Training Agency for International Development, U.S. Department of State. Washington, DC, Author, 1971.

American Language Institute (Georgetown). *A report on the results of English testing during the 1966 Pre-University Workshop at the American Language Institute*. Unpublished manuscript. Georgetown University, 1966.

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Psychological Association, 1985.

Angelis, P. J., Swinton, S. S., and Cowell, W. R. *The performance of non-native speakers of English on TOEFL and verbal aptitude tests* (TOEFL Research Report 3). Princeton, NJ: Educational Testing Service, 1979.

Angoff, W. A., and Sharon, A. T. *A comparison of scores earned on the Test of English as a Foreign Language by native American college students and foreign applicants to United States colleges* (Research Bulletin 70-8). Princeton, NJ: Educational Testing Service, 1970.

Chase, C. I., and Stallings, W. M. *Tests of English language as predictors of success for foreign students*. Indiana Studies in Prediction No. 8. Monograph of the Bureau of Educational Studies and Testing. Bloomington, IN: Bureau of Educational Studies and Testing, Indiana University, 1966.

Clark, J. L. D. *The performance of native speakers of English on the Test of English as a Foreign Language* (TOEFL Research Report 1). Princeton, NJ: Educational Testing Service, 1977.

Cowell, W. R. Item-response theory pre-equating in the TOEFL testing program. In P. W. Holland and D. B. Rubin (Eds.), *Test equating* (pp. 149-161). New York: Academic Press, 1982.

Duran, R. P., Canale, M., Penfield, J. Stansfield, C. W., and Liskin-Gasparro, J. *TOEFL from a communicative viewpoint on language proficiency: A working paper* (TOEFL Research Report 17). Princeton, NJ: Educational Testing Service, 1985.

Gershman, J. *Testing English as a foreign language: Michigan/TOEFL study*. Unpublished manuscript. Toronto Board of Education, 1977.

Heil, D. K., and Aleamoni, L. M. *Assessment of the proficiency in the use and understanding of English by foreign students as measured by the Test of English as a Foreign Language* (Report No. RR-350). Urbana, IL: University of Illinois. (ERIC Document Reproduction Service No. ED 093 948), 1974.

Homburg, T. J. TOEFL and GPA: an analysis of correlations. In R. Silverstein (Ed.), *Proceedings of the Third International Conference on Frontiers in Language Proficiency and Dominance Testing*. Occasional Papers on Linguistics, No. 6. Carbondale, IL: Southern Illinois University, 1979.

Hwang, K. Y., and Dizney, H. F. Predictive validity of the Test of English as a Foreign Language for Chinese graduate students at an American university. *Educational and Psychological Measurement*, 1970, *30*, 475-477.

Linn, R., and Slinde, J. The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, 1977, 47(1), 121-150.

Livingston, S. A., and Zieky, M. J. *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service, 1982.

Lord, F. M., and Novick, M. R. *Statistical theories of mental test scores*. Reading, MA: Addison Wesley, 1968.

Lord, F. M. *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum, 1980.

Maxwell, A. *A comparison of two English as a foreign language tests*. Unpublished manuscript. University of California (Davis), 1965.

McKinley, R. L. *An introduction to item response theory. Measurement and Evaluation in Counseling and Development*, 1989, *22* (1), 37-57.

Odunze, O. J. Test of English as a Foreign Language and first year GPA of Nigerian students (Doctoral dissertation, University of Missouri-Columbia, 1980). *Dissertation Abstracts International, 42*, 3419A-3420A. (University Microfilms No. 8202657), 1982.

Pack, A. C. A comparison between TOEFL and Michigan Test scores and student success in (1) freshman English and (2) completing a college program. *TESL Reporter*, 1972, *5*, 1-7, 9.

Pike, L. *An evaluation of alternative item formats for testing English as a foreign language* (TOEFL Research Report 2). Princeton, NJ: Educational Testing Service, 1976.

Pitcher, B., and Ra, J. B. *The relation between scores on the Test of English as a Foreign Language and ratings of actual theme writing* (Statistical Report 67-9). Princeton, NJ: Educational Testing Service, 1967.

Powers, D. E. *The relationship between scores on the Graduate Management Admission Test and the Test of English as a Foreign Language* (TOEFL Research Report 5). Princeton, NJ: Educational Testing Service, 1980.

Powers, D. E. *A survey of academic demands related to listening skills* (TOEFL Research Report 20). Princeton, NJ: Educational Testing Service, 1985.

Schrader, W. B., and Pitcher, B. *Interpreting performance of foreign law students on the Law School Admission Test and the Test of English as a Foreign Language* (Statistical Report 70-25). Princeton, NJ: Educational Testing Service, 1970.

Stansfield, C. W., Ed. *Toward communicative competence testing: Proceedings of the second TOEFL Invitational Conference* (TOEFL Research Report 21). Princeton, NJ: Educational Testing Service, 1986.

Swineford, F. Test analysis — Test of English as a Foreign Language (Statistical Report 71-112). Princeton, NJ: Educational Testing Service, 1971.

Swinton, S. S. *A manual for assessing language growth in instructional settings* (TOEFL Research Report 14). Princeton, NJ: Educational Testing Service, 1983.

Test of English as a Foreign Language: Interpretive Information. Princeton, NJ: College Entrance Examination Board and Educational Testing Service, 1970.

Thorndike, R. I., and Hagen, E. P. *Measurement and Evaluation in Education* (4th Ed.). New York: Wiley & Sons, Inc., 1977.

Upshur, J. A. *Comparison of performance on "Test of English as a Foreign Language" and "Michigan Test of English Language Proficiency."* Unpublished manuscript. University of Michigan, 1966.

Wilson, K. M. *GMAT and GRE Aptitude Test performance in relation to primary language and scores on TOEFL* (TOEFL Research Report 12). Princeton, NJ: Educational Testing Service, 1982.

ETS Field Service Offices

Bay Area Office
P.O. Box 23060
Oakland, CA 94623-2306
415-596-5500

Pasadena Office
Suite 540
2 North Lake
Pasadena, CA 91101-9852
818-578-1971

Midwestern Regional Office
Suite 300
One Rotary Center
1560 Sherman Avenue
Evanston, IL 60201-4896
708-869-7700

New England Office
Ten Brookline Place West
Suite 301
Brookline, MA 02146-7215
617-739-2210

Princeton Office
TOEFL/TSE Services
P.O. Box 6151
Princeton, NJ 08541-6151
215-750-8050

Puerto Rico Office
Metmor Plaza 11th Floor Suite 1110
268 Ponce de Leon Ave.
Hato Rey, PR 00918
809-753-6363

Southern Regional Office
Suite 1240
250 Piedmont Avenue, NE
Atlanta, GA 30308-3336
404-524-4501

Southwestern Regional Office
Austin Centre Building
701 Brazos Street, Suite 350
Austin, TX 78701-3262
512-478-8191

Washington, DC, Office
Suite 475
1825 Eye Street, NW
Washington, DC 20006-5403
202-659-0616

BEST COPY AVAILABLE

TOEFL Representatives

ALGERIA, BAHRAIN, IRAQ, KUWAIT, OMAN, QATAR, SAUDI ARABIA, SUDAN, UNITED ARAB EMIRATES:
AMIDEAST
Testing Programs
1100 17th Street, NW
Suite 300
Washington, DC 20036-4601, USA

BRASIL:
Instituto Brasil-Estados Unidos
Av. Nossa Senhora de Copacabana, 690-8° Andar
22050 Rio de Janeiro, RJ, Brasil

CANADA:
Information Centre for Canada
P.O. Box 162, Station S
Toronto, ON M5M 4L7
Canada

EGYPT:
AMIDEAST
6 Kamel El Shennawy Street
Second Floor
Apartment 5
Garden City, Cairo, Egypt

EUROPE:
(all countries, including Cyprus Great Britain, Iceland, and Turkey)
CITO-TOEFL
P.O. Box 1203
6801 BE Arnhem
Netherlands

HONG KONG:
Hong Kong Examinations Authority
San Po Kong Sub-office
17 Tseuk Luk Street
San Po Kong
Kowloon, Hong Kong

INDIA:
Institute of Psychological and Educational Measurement
25-A Mahatma Gandhi Marg
Allahabad, U.P. 211 001, India

INDONESIA:
Institute of International Education
P.O. Box 18 KBYCO
Jakarta Selatan 12951, Indonesia

JAPAN:
Council on International Educational Exchange
Hirakawa-cho Kaisaka Bldg. 1 F
1-6-8 Hirakawa-cho, Chiyoda-ku,
Tokyo, 102 Japan

JORDAN:
AMIDEAST
P.O. Box 1249
Amman, Jordan

KOREA:
Korean-American Educational Commission
K.P.O. Box 643
Seoul 110-606, Korea

LEBANON:
AMIDEAST
P.O. Box 135-155
Ras Beirut, Lebanon
or
AMIDEAST
P.O. Box 70-744
Antelias, Beirut, Lebanon

MALAYSIA:
MACEE
TOEFL/TSE Services
355, Jalan Ampang
50450 Kuala Lumpur, Malaysia

MEXICO:
Institute of International Education
Londres No. 16, 2nd Floor
Apartado Postal 61-115
Mexico 06600 D.F., Mexico

MOROCCO:
AMIDEAST
25 bis, Patrice Lumumba
Apt. No. 8
Rabat, Morocco

PEOPLE'S REPUBLIC OF CHINA:
China International Examinations Coordination Bureau
#30 Yuquan Road, 100039
Beijing
People's Republic of China

SINGAPORE:
MACEE
TOEFL/TSE Services
355, Jalan Ampang
50450 Kuala Lumpur, Malaysia

SYRIA:
AMIDEAST
P.O. Box 2313
Damascus, Syria

TAIWAN:
The Language Training & Testing Center
P.O. Box 23-41
Taipei 10098, Taiwan

THAILAND:
Institute of International Education
Room 21 9
A.U.A. Language Center
179 Rajadamri Road
G.P.O. Box 2050
Bangkok 10501, Thailand

TUNISIA:
AMIDEAST
BP 1134
Tunis, Tunisia

YEMEN ARAB REPUBLIC:
AMIDEAST
P.O. Box 1088
Sana'a, Yemen Arab Republic

Test of English as a Foreign Language
P.O. Box 6155
Princeton, NJ 08541-6155
USA

44

57516-03047 • S40P60 • 6780B8 • Printed in U.S.A.

BEST COPY AVAILABLE